

RAND **Statewide CalWORKs Evaluation**

Welfare Reform in California: Design of the Impact Analysis

*Jacob Alex Klerman, V. Joseph Hotz, Guido Imbens,
Paul Steinberg, Elaine Reardon, Patricia A. Ebener,
Jennifer A. Hawes-Dawson*

MR-1266.0-CDSS

July 2000

Prepared for the California Department of Social Services

L A B O R A N D P O P U L A T I O N

This is a final report of a project. It has been formally reviewed but has not been formally edited.

*RAND is a nonprofit institution that helps improve policy and decisionmaking through research and analysis.
RAND's publications do not necessarily reflect the opinions or policies of its research sponsors.*

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

20001211 110

PREFACE

In response to national welfare reform legislation--the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA), which was signed in August 1996--California passed legislation on August 11, 1997, that replaced the existing Aid to Families with Dependent Children (AFDC) and Greater Avenues to Independence (GAIN) programs with the California Work Opportunity and Responsibility to Kids (CalWORKs) program. Following an open and competitive bidding process, the California Department of Social Services (CDSS), which administers CalWORKs, awarded a contract to RAND to conduct a statewide evaluation of the CalWORKs program. That evaluation included both a process analysis examining how CalWORKs is being implemented and an impact analysis examining its costs and benefits.

This report presents an overview of RAND's plan for conducting the impact analysis component of the CalWORKs evaluation as of September 1999. Another document, MR-1266.0/1-CDSS, *Welfare Reform in California: Design of the Impact Analysis: Preliminary Investigations of Caseload Data*, Steven Haider, Jacob Alex Klerman, Jan M. Hanley, Laurie McDonald, Elizabeth Roth, Liisa Hiatt, and Marika Suttorp, discusses preliminary results and planned future analyses of caseload data.

For more information about the evaluation, see:

<http://www.rand.org/CalWORKs> or contact:

Jacob Alex Klerman
RAND
1700 Main Street
P.O. Box 2138
Santa Monica, CA 90407-2138
(310) 393-0411 x6289
klerman@rand.org

Aris St. James
CDSS
744 P Street, MS 12-56
Sacramento, CA 98514
(916) 657-1959
astjames@dss.ca.gov

CONTENTS

Preface.....	iii
Figures.....	vii
List of Tables.....	ix
Summary.....	xi
Introduction.....	xi
Impact Analysis Phases.....	xi
Phase 1: Describing Outcomes Under CalWORKs	xi
Phase 2: Estimating Causal Effects of Reform	xii
Phase 3: Analyzing Costs and Benefits	xiii
Approaches to Address Methodological Challenges.....	xiii
Data and Outcomes.....	xiv
Status.....	xv
Acknowledgments.....	xvii
List of Abbreviations.....	xix
1. Introduction.....	1
Background.....	1
Objectives.....	2
Organization of this Document.....	3
2. The Goals of the Impact Analysis.....	5
The Outcomes and Populations of Interest.....	5
Phase 1: Describing Outcomes under CalWORKs.....	8
Phase 2: Establishing Causal Effects of Reform.....	10
Baseline 1: Compared to Other States	11
Baseline 2: Compared to AFDC/Greater Avenues to Independence	12
Baseline 3: Compared to Another California County	13
Phase 3: Analyzing Costs and Benefits.....	14
3. Methodological Approach for Estimating Causal Effects.....	17
The Challenge of Estimating Causal Effects: The Problem of	
Confounders	17
Randomization and Confounders.....	22
The Need for Nonexperimental Approaches To Estimating Causal	
Effects	24
Simple Mean Difference Estimator Approach	25
Conventional Linear Regression Approach	27
The Statistical Matching Approach	30
Strategies for (Partially) Validating the Methods.....	31
Other Technical Issues in Using the Two Nonexperimental Approaches	
.....	34
Additional Data Requirements	34
Stratification	35
The Form of the Statistical Model	36
4. Data and Outcomes.....	37

Data Set Characteristics.....	37
The CPS: The Primary Data Set for Interstate Comparisons.....	39
The 6CHS: The Primary Data Set for Child and Family Outcomes...	41
Welfare System Outcomes.....	42
Caseloads	42
Caseload Dynamics	44
Aid Payments	46
Program Activities	47
Self-Sufficiency Outcomes.....	48
Employment and Earnings of Current and Past Recipients	49
Employment and Earnings of Potential Future Recipients	50
Hours of Work and Wages	52
Child and Family Well-Being Outcomes.....	52
Household Resources and Poverty	53
Marital Status	54
Births, Their Marital Context, and Child Health	55
Health Insurance	55
Foster Care, Child Abuse, and Child Living Arrangements	56
Financial Outcomes.....	56
Cash Aid Payments	57
County Administrative Expenditures	57
Other Data Sources for Financial Outcomes	58
5. The Six County Household Survey.....	61
Overview of Design.....	61
Limitations of the Design.....	65
Survey Content.....	66
6. Project Focus and Status.....	69
Focus of Analyses.....	69
1. Caseload, Aid Payments, and Costs	69
2. Program Activities	71
3. Employment and Earnings	71
4. Child and Family Well-Being	72
Status.....	73
References.....	77

FIGURES

Figure 3.1--The Core Evaluation Problem	18
Figure 3.2--National and California Caseloads	20
Figure 3.3--National and California Unemployment Rates	20
Figure 3.4--California Regional Caseloads	21
Figure 3.5--California Regional Unemployment Rates	21

LIST OF TABLES

Table S.1--Uses of Various Data Sources in Relation to Outcomes of Interest	xv
Table 4.1--Characteristics of Primary Data Sources.....	38
Table 4.2--Use of Various Data Sources--Welfare System Outcomes.....	42
Table 4.3--Use Of Various Data Sources--Self-Sufficiency Outcomes....	49
Table 4.4--Use of Various Data Sources--Child and Family Well-Being Outcomes	53
Table 5.1--Key Design Features of the 6CHS.....	62
Table 5.2--Tabulations for Sample Characteristics.....	64
Table 5.3--Sections of the Survey.....	67

SUMMARY

INTRODUCTION

California's response to the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (PRWORA) was the California Work Opportunity and Responsibility to Kids (CalWORKs) program--a "work first" program that provides support services to help recipients move from welfare to work and toward self-sufficiency. The California Department of Social Services (CDSS)--the state agency in charge of welfare--contracted with RAND for an independent evaluation of CalWORKs to assess both the policy implementation and its impact, at both the state and county levels. RAND is now working on the first phase of the impact analysis component of the evaluation, the results of which are scheduled for release in October 2000. The final impact analysis report is due to be released in October 2001.

This report presents a detailed plan for how RAND will conduct the impact analysis. The report discusses the three phases of the impact analysis: (1) describing outcomes under CalWORKs; (2) establishing the causal effects of reform; and (3) analyzing costs and benefits. It also reviews the outcomes of interest: welfare system outcomes; self-sufficiency and employment outcomes; family and child well-being outcomes; and financial outcomes. Finally, it examines the methodological challenges involved in conducting the analysis and our proposed solutions, as well as the data sets that will be used to conduct the analysis, their limitations, and our solutions for dealing with those limitations.

IMPACT ANALYSIS PHASES

Phase 1: Describing Outcomes Under CalWORKs

The first phase of the analysis--describing outcomes under CalWORKs--is important in its own right and crucial for the two following phases. Some outcomes can be judged against objective standards: Are county CalWORKs programs meeting participation rate requirements? For which subgroups? What portion of current recipients

is working? What portion of current recipients is in poverty? What portion of recent recipients is in poverty? How does that portion vary with time since leaving aid? Our ability to conduct this phase of the impact analysis depends on the availability of appropriate data.

Phase 2: Estimating Causal Effects of Reform

The second phase of the analysis will attempt to estimate the effects of CalWORKs on the outcomes of interest relative to various alternative programs or environments. We have identified three such alternatives (called baselines or counterfactuals):

- (1) **Compared to Other States.** Every state and many other governmental units are reforming their welfare programs to be consistent with PRWORA and to exploit the new latitude that PRWORA provides. Ideally, we would like to know what California outcomes *would have been* if California had adopted the PRWORA plan of some other state. To do so, we would compare its outcomes to outcomes of other states. If, *holding all else equal*, other states have considerably better outcomes, California might consider modifying CalWORKs to resemble aspects of welfare programs in those states more closely.
- (2) **Compared to AFDC/Greater Avenues to Independence (GAIN).** Before and after comparisons are natural for the evaluation of a new program. CalWORKs replaces AFDC/GAIN, which means a natural comparison is to what the outcomes *would have been* if AFDC/GAIN had been left in place. This perspective is useful in evaluating PRWORA and CalWORKs.
- (3) **Compared to Another California County.** Just as PRWORA gave the states increased latitude in designing their post-reform welfare programs, CalWORKs also gave California's counties increased latitude. We expect the implementation of CalWORKs to vary considerably across the counties. We will use this variation in county welfare programs to attempt to explain variation in outcomes across counties. Even without change in the CalWORKs legislation, individual counties can use the

results of such comparisons to fine-tune or revamp their welfare programs.

Phase 3: Analyzing Costs and Benefits

In addition to describing the outcomes of interest, the RFP requested a cost-benefit analysis of those outcomes. Because of the way we receive the data, we find it more helpful to partition not according to "benefits" and "costs" but according to "effects on government finances" (e.g., direct costs of welfare programs, such as direct cash payments for benefits, and indirect costs, such as increased tax revenues and workers' compensation payments) and "non-financial effects" (e.g., changes in the number and characteristics of individuals receiving cash aid and other welfare programs).

APPROACHES TO ADDRESS METHODOLOGICAL CHALLENGES

The process of estimating causal effects--the intent of phase 2--is considerably more difficult than the process of describing outcomes in phase 1. Estimating such causal effects requires being able to isolate the pure effect of the CalWORKs legislation (or of the CalWORKs program of a given county), which means we need to control for the effects of the other things that vary across time and place--referred to as confounders. This is the methodological challenge. Random assignment, a relatively assumption-free approach to this challenge, was not feasible given the dramatic change in the welfare system under CalWORKs, which was designed not just to reform a bureaucracy but to change public attitudes about welfare.

Instead, we will apply best practices from the nonexperimental evaluation literature. These best practices include difference-of-differences regression and statistical matching.

While such nonexperimental evaluation approaches are promising, they rely on untestable assumptions that are rarely exactly applicable. Other independent analysts will sometimes reach different conclusions. We will highlight where we have great confidence in our methods and where we need to be more cautious; and those using the results of our evaluation should consider the resulting uncertainty when reviewing and applying the results.

DATA AND OUTCOMES

There is a close connection between the characteristics of data sets and their usability for conducting analyses of the effects of CalWORKs. Ideally, for each outcome of interest, we would have data for both before and after CalWORKs, for every person (not merely a sample), for each of California's 58 counties (and ideally all 50 states), and for current, former, and potential future recipients. Of course, no data set is ideal on each of these criteria.

The major primary data available for conducting the impact analysis are state and county welfare administrative data systems and information on earnings from unemployment insurance and tax filings. However, these data are insufficient to address all the outcomes of interest. In particular, information on child and family outcomes for current recipients is poor and is worse for former or potential recipients. To compensate for these weaknesses in the available administrative data, RAND is fielding the Six-County Household Survey (6CHS) within the six focus counties specified by CDSS: Alameda, Butte, Fresno, Los Angeles, Sacramento, and San Diego. The 6CHS will interview current and recent recipients in each of the six focus counties.

In addition, to be able to do the interstate descriptive analyses and to estimate causal effects using other states as a baseline, we need an ongoing national, general-purpose survey. We have chosen to work with the U.S. Bureau of the Census's Demographic Supplement to the March Current Population Survey (CPS). Because of its national coverage, the CPS will be used by many other analysts across the nation. Using the CPS data, we will be able to reexamine those national analyses from a California perspective (e.g., to determine the implied outcome in California, given the outcomes in other states).

Table S.1 summarizes the data sources in terms of the key elements within the three nonfinancial outcomes of interest. Coverage varies across data sets: some have information only on current recipients, others have information about a broader population.

Table S.1

Uses of Various Data Sources in Relation to Outcomes of Interest

Specific Outcomes/Elements	CPS	6CHS	MEDS	Q5	6CWAD	MEDS- EDD
WELFARE SYSTEM						
Caseload	X	X	X	X	X	
Caseload Dynamics				X	X	
Aid Payments				X	X	
Program activities				X	X	
SELF-SUFFICIENCY						
Employment and Earnings of Current Recipients	X	X			X	X
Employment and Earnings of Past Recipients		X				X
Employment and Earnings of Potential Future Recipients	X					
Hours of Work and Wages	X	X		X	X	
CHILD AND FAMILY WELL-BEING						
Household Resources and Poverty	X	X		X	X	
Marital Status	X	X		X	X	
Births, Their Marital Context, and Child Health	X	X				
Health Insurance	X	X	X	X	X	X
Foster Care, Child Abuse, and Child Living Arrangements	X	X				

Abbreviations: CPS = Current Population Survey; 6CHS = Six County Household Survey; MEDS = Medi-Cal Eligibility Determination System; Q5 = Quality Control data; 6CWAD = Six County Welfare Administrative Data systems; MEDS-EDD = MEDS-Employment Development Department earnings match; EDD = Employment Development Department earnings match.

Notes: X = The data contain this element (subject to quality assessment).

To conduct the cost-benefit analysis, we will draw on budget information describing expenditures that flow from CDSS to the counties, looking at cash aid payments, county administrative expenditures, and other financial data sources.

STATUS

Our analysis plan will evolve as we learn more about the data and as preliminary results emerge. We expect our plans for the impact analysis to continue to evolve over the remaining two years of the evaluation. Future quarterly progress reports, meetings of the Advisory Committee, draft documents, and presentations of plans and results

before academic and policy audiences will provide opportunities for RAND to share these evolving plans with CDSS and the broader research community. Feedback from future written and oral presentations will also help RAND improve the technical quality of its analyses and the allocation of available resources to the tasks of greatest interest to CDSS.

ACKNOWLEDGMENTS

This report is based on the work of the team of programmers led by Jan Hanley and including Debbie Wesley, Beth Roth, Laurie McDonald, Shaoling Zhu and Rodger Madison, as well as on the work of several research assistants: Lee Mizell, Liisa Hiatt, and Marika Suttorp.

This report also draws on the efforts of the RAND CalWORKs team, including the co-principal investigator Gail Zellman and the other senior staff members for the process analysis team: Nicole Humphrey, Tammi Chun, and Patricia Ebener. We also wish to acknowledge Patrice Lester, Natasha Kostan, Joan Verdon, and Christopher Dirks, who prepared the manuscript; Betty Amo, who expedited the publication process; and Phyllis Gilmore, who did an outstanding job of editing.

Among the several internal and external reviewers who provided valuable comments, we would like to acknowledge Steven Haider and Elaine Reardon. Beyond those attending the Advisory Committee and the Technical Subcommittee meetings, Werner Schink and Paul Smilanick provided especially useful observations.

LIST OF ABBREVIATIONS

Symbol	Definition
6CHS	RAND's Six-County Household Survey
6CWAD	Six-County Welfare Administrative Data
AB	Assembly Bill
ACIS	RAND All-County Implementation Survey
ACL	All-County Letter
AFDC	Aid to Families with Dependent Children
APP	Alternative payment provider
BC	Birth Certificate
BoS	Board of Supervisors
CalWORKs	California Work Opportunity and Responsibility to Kids
CBO	Community-based organization
CDE	California Department of Education
CDSS	California Department of Social Services
COLA	Cost-of-living adjustment
CSAC	California State Association of Counties
CWDA	California Welfare Directors' Association
CPS	Current Population Survey
DHHS	U.S. Department of Health and Human Services
DoL	Department of Labor
EITC	Earned Income Tax Credit
EDD	Employment Development Department
EW	Eligibility worker
FG	Family Grant assistance unit
FSA	Family Support Act of 1988
GAIN	Greater Avenues to Independence
GAO	General Accounting Office
H.R.	House Resolution
JOBS	Job Opportunities and Basic Skills (training program)
JTPA	Job Training Partnership Act
MDRC	Manpower Demonstration Research Corporation
MEDS	Medi-Cal Eligibility Determination System
MEDS-EDD	MEDS-Employment Development Department earnings match
MOE	Maintenance of effort (requirement)
NICHD	National Institute of Child Health and Human Development
OJT	On-the-job training
PG	Parent Group
PIC	Private Industry Council
PRWORA	Personal Responsibility and Work Opportunity Reconciliation Act
Q5	Quality control data
REB	CDSS Research and Evaluation Branch
RFP	Request for proposal
SB	Senate bill
SEC	Section
SIP	Self-initiated program

SIPP	Survey of Income and Program Participation
SPD	Survey of Program Dynamics
SRG	RAND's Survey Research Group
SSI	Supplemental Security Income
TANF	Temporary Assistance to Needy Families
UP	Unemployed Parent assistance unit
WIC	California Welfare and Institutions Code
WIN	Work Incentive program
Work Pays	California's Assistance Payments and Work Pays Demonstration Project
WTW	Welfare-to-Work

1. INTRODUCTION

BACKGROUND

The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (PRWORA) fundamentally changed the American welfare system, replacing the Aid to Families with Dependent Children (AFDC) program with the Temporary Assistance for Needy Families (TANF) program. In addition, PRWORA deliberately and decisively shifted the authority to shape welfare programs from the federal government to the individual states. California's response to PRWORA was the California Work Opportunity and Responsibility to Kids (CalWORKs) program--a "work first" program that provides support services to help recipients move from welfare to work and toward self-sufficiency. Beyond encouraging the transitions to work and self-sufficiency, CalWORKs also imposes lifetime time limits to further motivate recipients to make these transitions. Finally, CalWORKs devolves much of the responsibility and authority for implementation to California's 58 counties, increasing counties' flexibility and financial accountability in designing their welfare programs.

The California Department of Social Services (CDSS)--the state agency in charge of welfare--contracted with RAND for an independent evaluation of CalWORKs to assess both the process (or implementation) and its impact (or outcomes), at both the state and county levels. RAND has released the findings of the first phase of the process analysis in a series of documents¹; two follow-on process-analysis reports for the subsequent two phases are due to be released in February 2000 and February 2001.

RAND is now working on the first phase of the impact-analysis component of the evaluation, the results of which are scheduled for release in October 2000. The final impact-analysis report is due to be released in October 2001. The original request for proposal (RFP)

¹ See Zellman et al., (1999a, 1999b); Ebener and Klerman (1999); and Ebener, Roth, and Klerman (1999).

(CDSS, 1998, p. 5) called for a statewide analysis of outcomes and of costs and benefits, as well as for similar analyses of six focus counties (CDSS, 1998, p. 7): Alameda, Butte, Fresno, Los Angeles, Sacramento, and San Diego.

In terms of the outcomes of interest, we will study the welfare system (e.g., how welfare recipients are flowing through the mandated CalWORKs welfare-to-work [WTW] activities); the transition to self-sufficiency (e.g., the effects on employment, earnings and hours worked), and child and family well-being (e.g., the effects on child poverty rate).

The costs and benefits the impact analysis will consider are direct payments made to families, payments made to service providers (including transportation, child care and other supplementary services), indirect costs and revenues, including increased income tax payments, and the administrative costs of the state and county welfare agencies operating the CalWORKs program.

Where possible, we will analyze these outcomes statewide using administrative data, augmenting our analyses with data on California residents collected as part of nationally representative surveys. Many outcomes of interest, however, are not measured by these data. To allow exploration of these otherwise unmeasured outcomes, we will devote considerable resources to transforming county-level administrative data into analysis files in the six focus counties. We will also collect primary data through a household survey--RAND's Six-County Household Survey (6CHS) (described in more detail in Section 5)--to obtain information about outcomes that have not been recorded in administrative records, designed as they were for record keeping under the old AFDC system.

OBJECTIVES

This report presents a more detailed plan for how RAND will conduct the impact analysis. While some uncertainty about data systems still remains, we have made considerable progress in this area. In addition, we have gained a better understanding of the CalWORKs program from the first phase of the process analysis. This improved understanding has

affected our plans for the impact analysis. Finally, given the additional time available, we have devoted more thought to some of the more difficult analytic issues. Taken together, we now have the ability to provide a considerably more detailed analysis plan for the main data systems than what appeared in our original proposal.

ORGANIZATION OF THIS DOCUMENT

The next section of the report lays out in broad strokes what we intend to accomplish in the impact analysis, focusing on three phases of the impact analysis--(1) describing outcomes under CalWORKs; (2) establishing the causal effects of reform; and (3) analyzing costs and benefits. While the methodological approach for doing the descriptive analyses in phase 1 is fairly clear-cut, the approach needed for the causal analyses in phases 2 and 3 are more complicated. Thus, Section 3 describes the methodological approach we will use in the phase 2 and 3 effort in more detail. Section 4 discusses each outcome and the data systems we will employ across all phases of the analyses. Section 5 discusses our household survey effort--the 6CHS--in more detail. Section 6 concludes with a discussion of the project's focus and our current status.

A complementary RAND report documents in more detail the statewide data systems on welfare participation and provides preliminary results: See Haider et al., MR-1266.0/1-CDSS, 1999.

2. THE GOALS OF THE IMPACT ANALYSIS

As mentioned in Section 1, to meet the goals of the impact analysis, we plan a three phase effort: (1) describing outcomes under CalWORKs, (2) estimating the causal effects of reform, and (3) analyzing costs and benefits.

This section begins with a review of the outcomes of interest as described in CDSS's RFP and then provides a brief overview of the general issues in each of the three phases of the analysis.

THE OUTCOMES AND POPULATIONS OF INTEREST

In its discussion of the "Statewide Impact and Cost-Benefit Study" component, CDSS's RFP (CDSS, 1998) for the evaluation describes the outcomes of interest as follows:

On a statewide basis, what is the impact of CalWORKs on:

- The incidence of aid receipt including Food Stamps, cash aid and Medi-Cal, SSI, employment, and earnings of current and former CalWORKs recipients?
- Family structure, including the number of two-parent families that become one-parent households or vice versa, and the movement of children into and out of the household in current and former CalWORKs households, including movement into and out of foster care?
- The well being of children, including entries into foster care, rates of child poverty, and frequency of at-risk births and child abuse among current and former CalWORKs recipients?
- What are the costs and benefits of the CalWORKs program?

Costs and benefits should include those that have been measured, that are measurable, but have not been measured, and those that are intangible and not subject to measurement.

The "County-Level Impact and Cost-Benefit Study" includes all the statewide outcomes and adds:

- Local Government?
- What function do Private Industry Councils and child care and child support services play in positive impacts of CalWORKs? What part do these factors play in the negative impacts?

We find it useful to organize our thinking about the outcomes specified in the RFP as follows. The evaluation needs to be sensitive to the objectives of welfare reform. PRWORA was intended to actively move current welfare recipients off welfare to self-sufficiency and to discourage potential future welfare recipients from life choices that are likely to cause them to become welfare recipients. This organization of the outcomes suggests that the evaluation needs to consider impacts not only on current recipients but also on former recipients and on potential future recipients. This need, in turn, influences the comparisons that will be made across counties and states and over time.

Outcomes for *current recipients* are the easiest to monitor. These outcomes are relatively well-recorded in the administrative records of the welfare program. However, in considering impacts on current recipients, the evaluation needs to remember that a finding about effects on current recipients is difficult to interpret in isolation. For example, results might show that those who remain on aid are worse off under CalWORKs but that those who left the aid rolls are much better off; an evaluation that focused only on active aid recipients would miss the improved life circumstances of former recipients. Thus, to evaluate a policy, we must consider both the effects on current recipients and the effects on recent recipients.

A primary goal of CalWORKs is to move recipients promptly to work and self-sufficiency. To assess how well this goal is achieved, we need to know what share of the caseload at a point in time is no longer receiving cash assistance. Similarly, among those who no longer receive cash aid (i.e., CalWORKs), we need to know what happens to them. For many purposes, the appropriate comparison averages over people who are still recipients and those who no longer are recipients. For example,

in comparing outcomes in two counties, we might start with two groups of people--one in each county--each of whom received cash assistance in some earlier calendar month. We would then compare subsequent outcomes across these two groups, regardless of whether they are still aid recipients.

Arguably, just as important as moving current recipients to self-sufficiency, CalWORKs aims to discourage people from ever going on the welfare rolls or, if they do enter, from remaining for long periods of time (i.e., becoming dependent on welfare). The logic is that, if people know that cash assistance is strictly time-limited (five years) and will require work, they have greater incentives to take actions that will make them and their families more self-sufficient. For example, they will develop labor-market skills (e.g., finish high school), delay parenthood until finishing school, and marry before having children. Providing families with incentives to change their life-course strategies in ways that are likely to reduce their entry into welfare figured prominently in the welfare reform debate and has the potential for the most sweeping long-term effects. An evaluation of CalWORKs should attempt to measure such effects on *potential future welfare recipients*.²

Thus, to determine if this goal for CalWORKs is being achieved and to measure accurately one of its potential benefits, the evaluation should consider outcomes, not merely for current or recent recipients, but also for groups in the population who are not on welfare. In practice, we have strong reason to believe that these benefits will be concentrated in women of an age when they might, at least from the perspective of CalWORKs, "prematurely" have a first child. Moreover, we would expect to find stronger impacts of CalWORKs for such "at-risk" groups than for those segments of the population who are less likely to be on welfare.

² Of course, there were "potential recipients" in the pre-CalWORKs era as well. More generally, we want to measure the effect of CalWORKs on entry into cash aid receipt. Different policies will induce different people to come onto aid.

Consideration of effects on future beneficiaries has two implications for our evaluation. First, we need data on the outcomes of interest for samples drawn from a broader population than current or former AFDC/TANF recipients. Second, we do not want to define the at-risk population too broadly. Even a strong effect on a small group would be hard to detect if the data contain mostly people for whom welfare reform is essentially irrelevant (e.g., 45-year-old working men). Obtaining data for such samples for the full range of outcomes of interest presents a challenge. We discuss the nature of the challenge and possible solutions in the context of effects on earnings in Section 4, "Employment and Earnings of Current and Past Recipients."

PHASE 1: DESCRIBING OUTCOMES UNDER CALWORKS

The first phase of the analysis--describing outcomes under CalWORKs--is important in its own right and crucial for phase 2 and phase 3. However, such description is often of great use in formulating policy. Some outcomes can be judged against objective standards: Are county CalWORKs programs meeting participation rate requirements? For which subgroups? What portion of current recipients is working? What portion of current recipients is in poverty? What portion of recent recipients is in poverty? How does that portion vary with time since leaving aid? As we discuss in detail below, our ability to conduct this phase of the impact analysis depends on the availability of appropriate data.

We envision this description phase as rich and multifaceted. Consider, for example, caseloads--an outcome for which the data are nearly ideal. We will begin with the aggregate descriptions of the caseload. How has the caseload varied over time? From before CalWORKs to after CalWORKs? Since the passage of the CalWORKs legislation as the programs have matured? How do the trends in California compare with those in other states? How do the levels of participation and trends vary across California's counties? Carefully considering the timing and geographical patterns in caseload trends often provides insights for understanding the effects of the program.

These analyses concern aggregate caseload counts. For California, the Medi-Cal Eligibility Determination System (MEDS) data (discussed in detail in Section 4) provide information on recipients' aid code (family group [FG], unemployed parent [UP], foster care, other child only) and demographic characteristics (gender, race, ethnicity, language, age, number and age of children). These individual-level data allow us to consider the level and trends in the caseload separately for sub-groups. Such disaggregated results often yields insight into aggregate caseload trends and the effects of the program. For example, some have claimed that much of the decline in California's caseload is concentrated among Hispanics and results from changes in perceived policies and attitudes toward immigrants--legal and illegal--rather than from welfare reform.

Simple percentages can be informative but for many purposes, the appropriate concept is a rate: What is the probability that a given individual will receive aid? The appropriate rate is usually the ratio of the number of cases to the number of people from some population subgroup (e.g., blacks, Fresno County). Trends in these ratios are informative though a little trickier to interpret, because change over time could be driven by changes in the numerator or in the denominator. For example, the proper understanding of a program's effects on an observed caseload decline will vary depending on whether the probability that a young mother of a new child applies for aid is falling or whether, instead, the number of young women having first children is falling. Appropriate policy responses vary greatly depending on the relative importance of each component.

Moreover, the aggregate caseload at a point in time is the result of the earlier history of individual level decisions. Some people chose first to receive cash assistance in a given month, while some people chose not to do so. Among those who received cash assistance in previous months, some have received cash assistance continuously since first receipt. Others chose to leave cash assistance at some time, some of whom have already returned to cash receipt in earlier months.

Our description of outcomes under CalWORKs should consider such individual dynamics. What is the probability that an individual will first receive cash assistance in a given month? What is the probability

that this individual will stop receiving cash assistance in a given month? What is the probability that this individual will resume receiving cash assistance in a given month? How do these probabilities vary through time? Across the state's counties? Across program types? Across demographic subgroups? With the earlier history of receipt of cash assistance--age at first receipt, time since first receipt, total months of receipt, time in the current spell of receipt (i.e., period of continuous receipt), time since last receipt--we can better understand how these individual decisions affect the aggregate caseload trends and how these decisions are affected by CalWORKs. Some programs would be expected primarily to have deterrent effects (e.g., diversion), while others would be expected primarily affect individuals new to cash receipt (e.g., Job Club). Other program components can be examined the same way.

The descriptive phase of the impact analysis should and will consider each of these perspectives--aggregate, over time, by program type and by demographic group, total numbers and rates, static analyses and dynamic analyses.

The previous paragraphs have used caseloads as an illustrative example. As much as the data allow, we also expect to perform similar rich and multifaceted descriptive analyses of other outcomes, including aid payments, employment, earnings, child living arrangements, food security, and housing security.

PHASE 2: ESTABLISHING CAUSAL EFFECTS OF REFORM

The previous subsection discussed how we will describe outcomes under CalWORKs. The second phase of the analysis will attempt to estimate the effects of CalWORKs on the outcomes of interest relative to various alternative programs or environments. The process of estimating such causal effects is considerably more difficult than the process of describing outcomes. As we discuss in detail in Section 3, the standard, relatively assumption-free approach--random assignment--is not available. Instead, we need to apply best practices from the nonexperimental evaluation literature. These best practices include new methodological work being done as part of this evaluation. While such

nonexperimental evaluation approaches are promising, they do, nonetheless, rely on untestable assumptions that are rarely exactly applicable. Other independent analysts will sometimes reach different conclusions. Thus, we, the official evaluators, and those using the results of our evaluation methods will consider the resulting uncertainty when reviewing and applying the results.

Beyond issues about methods, estimating causal relations raises data issues. Some approaches require data from before and after CalWORKs, some approaches require data from other states in addition to California, and some approaches require data from many counties. In general, estimating causal effects will require more of everything: more observations, more counties, more time periods, and more control variables. In some cases the available data imply that, for some outcomes of interest, causal analyses may not be possible or will be estimated so imprecisely as to be of little use for policy evaluation. Section 4 has a more thorough discussion of all the data sets proposed for the analyses.

The causal effect of CalWORKs is defined as the observed outcomes under CalWORKs compared to what outcomes would have been under some baseline (sometimes called a *counterfactual*). We have identified three such baselines for the phase 2 analysis, which are described below, along with their general data requirements.

Baseline 1: Compared to Other States

Every state and many other governmental units are reforming their welfare programs to be consistent with PRWORA and to exploit the new latitude that PRWORA provides. Under this baseline, we would like to know what California outcomes *would have been* if California had adopted the PRWORA plan of some other state. To do so, we would compare its outcomes to outcomes of other states. If, *holding all else equal*, other states have considerably better outcomes, California might consider modifying CalWORKs to resemble aspects of welfare programs in those states more closely.

Estimating CalWORKs outcomes relative to what would have occurred if California had adopted a TANF program resembling that of some other

state requires consistent data across the states. Clearly, the California-specific data to which we have access as the official evaluators are of little use for such comparisons. In Section 4, we describe our primary data set for such interstate comparisons: the Current Population Survey (CPS). However, we do not anticipate allocating a significant portion of our resources to making these comparisons relative to the before-and-after cross-county analyses. Furthermore, such interstate comparisons are being done by several national evaluation efforts. We will review those studies and compare our results with other published analyses in the rest of the country.

Baseline 2: Compared to AFDC/Greater Avenues to Independence

CalWORKs replaces AFDC and Greater Avenues to Independence (GAIN), California's welfare-to-work program under AFDC. Thus, a natural comparison is to what the outcomes would have been if AFDC/GAIN had been left in place. This perspective is useful in evaluating PRWORA. It is also useful in evaluating CalWORKs. Before and after comparisons--AFDC/GAIN versus CalWORKs--are natural for the evaluation of a new program. What would outcomes have been if the old program had continued?

We note that this is a different question from the descriptive question we asked earlier: How have outcomes varied across time? In contrast, the causal estimate should hold constant everything but the shift from AFDC/GAIN to CalWORKs. In particular, we would project what outcomes would have been if AFDC/GAIN had continued but if everything else had continued to evolve as they have: labor-market conditions, exogenous changes³ in birth rates and marriage patterns, and other policy changes. This is a technically formidable task.

We also note that the direct usefulness of estimates of the effect of CalWORKs relative to this baseline is limited. Even if California concluded that outcomes would have been preferable under the AFDC/GAIN rules, the PRWORA funding rules would make it very expensive for the state to return to the AFDC/GAIN policy. However, this perspective does

³ By "exogenous changes," we mean changes not caused by welfare program changes.

provide insight into how extensively CalWORKs has changed various elements of the welfare system.

Estimating CalWORKs outcomes relative to what would have occurred if California had continued its AFDC/GAIN program requires consistent data from before and after the reforms. Such data are more readily available for some outcomes than for others. For other outcomes, such comparisons are simply not meaningful.

Baseline 3: Compared to Another California County

Just as PRWORA gave the states increased latitude in designing their post-reform welfare programs, CalWORKs also gave California's counties increased latitude. We expect the implementation of CalWORKs to vary considerably across the counties. We will use this variation in county welfare programs to attempt to explain variation in outcomes across counties. These comparisons are potentially quite useful. Even without change in the CalWORKs legislation, individual counties can use the results of such comparisons to fine-tune or revamp their welfare programs.

Again, this is a different question from the corresponding descriptive question: How do outcomes under CalWORKs vary across California's counties? Counties differ for a lot of reasons besides their CalWORKs programs. To isolate the effect of the program as opposed to these differences, the causal estimate should hold constant everything but the counties' CalWORKs programs. In effect, we would project what outcomes would have been if County A had adopted County B's CalWORKs program. As was true for the other two baselines, this is a technically formidable task that requires a lot of data about counties and how they differ. To be useful for this baseline, a data source must include consistent information for sufficiently large samples for at least two counties and ideally, for a large number of counties. Furthermore, to control for persistent inter-county differences, having consistent data for several years (in practice, from before and after CalWORKs) is nearly a prerequisite.

PHASE 3: ANALYZING COSTS AND BENEFITS

In addition to describing the outcomes of interest, the RFP requested a cost-benefit analysis of those outcomes. Because of the way we receive the data, we find it more helpful to partition not according to "benefits" and "costs" but according to "effects on government finances" and "non-government-financed effects."

The effects on government finances include the direct costs of welfare programs--direct cash payments for benefits, such as cash assistance, Food Stamps, and Medi-Cal; payments for services delivered, such as training, substance abuse treatment, and child care; and the administrative costs of running the program, which are expected to increase per recipient with the reform's more-intensive case management. In addition, there are indirect effects, including increased tax revenues (e.g., income taxes and payroll taxes) and spillover effects on other social programs (e.g., workers' compensation insurance, unemployment insurance and General Assistance). Some of these costs accrue to the federal government, some to the state government, and some to the county governments. In addition, some are true costs and some are negative "costs" (i.e., net income to government).

We will collect information on each of these financial effects at the federal, state, and county levels. Some of this information is available electronically; however, much of it is most easily obtained from official reports and in the administrative offices of federal, state, and county welfare agencies. Thus, we will collect some of these data as part of the preparation for the site visits and key informant interviews that are being conducted as part of the state and county process analysis, and the All-County Implementation Survey (ACIS). We will therefore have varying levels of detail across the different groups of counties.

The non-government-finance outcomes we will explore and the ones discussed for description (phase 1) and effects (phase 2) of reform include changes in the number and characteristics of individuals receiving cash aid and other welfare programs; changes in labor-market outcomes for current, recent, and potential future welfare participants; and changes in child and family outcomes, for current, recent, and

potential future welfare recipients and their children. Generically, we view these outcomes as the "benefits" of the reforms. Again, they enter the benefit-cost calculations as the causal effect of the reforms on outcomes.

For a benefit-cost computation, we want to compare net costs, conceptualized as effects on government budgets against net benefits (i.e., nonfinancial outcomes). We use net to refer to observed financial effects compared to what financial outcomes would have been under some alternative baseline. Again, doing so will require modeling costs under each system. In particular, such models need to consider how costs would have varied across caseloads that vary in absolute size and in their composition.

As was true for the phase 2 analysis, phase 3 also requires conducting causal analyses. Our methods for accomplishing causal analyses are the subject of the next section.

3. METHODOLOGICAL APPROACH FOR ESTIMATING CAUSAL EFFECTS

As discussed in Section 2, one crucial analytic goal of the evaluation is to estimate the *causal effects* of the CalWORKs legislation (i.e., to compare outcomes under CalWORKs to what the outcomes would have been under some alternative set of welfare rules or program implementations). This is the goal of the analyses to be conducted in phases 2 and 3.

In this section, we discuss the methodological challenge such causal analyses present, in particular, the problem of confounders and approaches to dealing with them. We then discuss the standard experimental approach to the methodological challenge and why the Statewide CalWORKs Evaluation must use other nonexperimental methods. We then discuss these methods. Finally, we discuss some technical issues concerned with using nonexperimental approaches.

THE CHALLENGE OF ESTIMATING CAUSAL EFFECTS: THE PROBLEM OF CONFOUNDERS

We can illustrate the methodological challenge--estimating causal effects in the presence of confounders--using the example of caseloads. Figure 3.1 plots caseloads over time. The solid line in the figure plots observed caseloads in California (from CA237 data). The dotted line diverges from the solid line in September 1997 (i.e., in the month following the passage of the CalWORKs legislation). This line is intended to represent what caseloads might have been if some specified other policy had been adopted. For this discussion, consider comparing outcomes under CalWORKs to what outcomes would have been if AFDC/GAIN had continued.⁴ If we knew what caseloads would have been under that alternative policy, the "impact" of CalWORKs on caseloads (relative to the baseline) would be the shaded area between the two lines.

⁴ This is the second baseline discussed in the previous section. Similar analyses apply to the other baselines discussed there--if California had adopted the PRWORA program of some other state or if one California county had adopted the welfare programs of some other county.

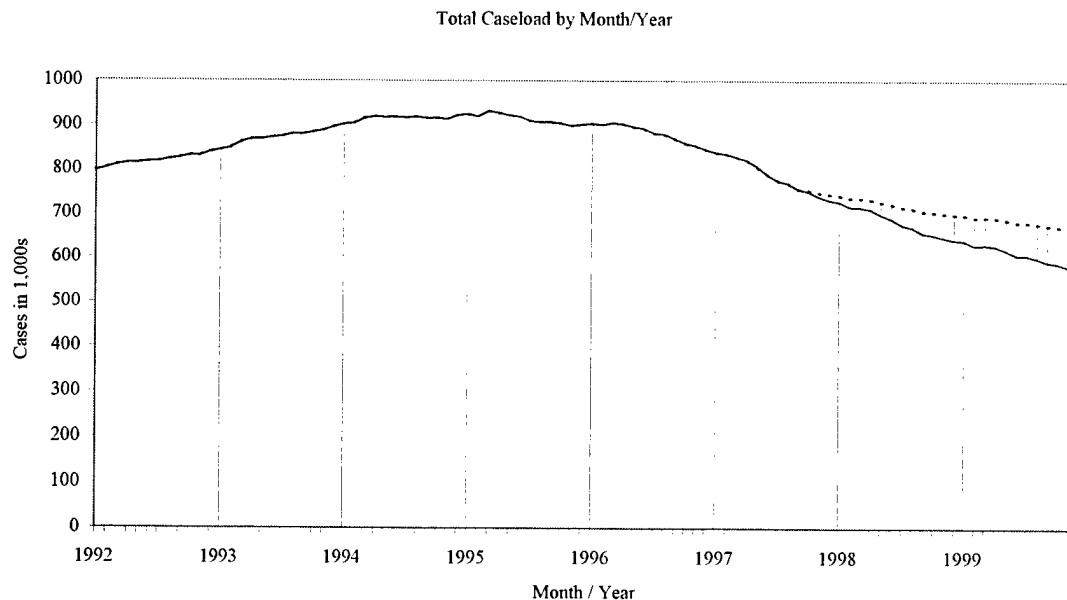


Figure 3.1--The Core Evaluation Problem

Of course, what makes estimating the impact of CalWORKs methodologically challenging is that we *do not observe* outcomes under the specified alternative policy. Instead, the evaluation team needs to *predict* what outcomes would have been under the specified alternative policy. Doing so is a nontrivial task.

One natural way to predict what outcomes would have been if AFDC/GAIN had continued would be to assume that in the absence of reform, caseloads would have continued at their level immediately before reform. Following this approach, we would conclude that CalWORKs has already had large effects on the caseload, because, since the passage of the legislation, the caseload has declined by 20 percent. This is the method of evaluation implicit in comparing current caseloads with those under AFDC/GAIN.

It is fairly easy to see what is wrong with this approach. We want to predict what outcomes would have been if AFDC/GAIN had continued *but if everything else had evolved as it has*. Even in the absence of changes in the welfare program, we would have expected caseloads to vary

through time. A cursory examination of the figure suggests that caseloads were falling prior to CalWORKs.⁵

Beyond welfare policy, why might the caseload change? Improved economic conditions is a prominent candidate. The nation as a whole, and California in particular, is in the midst of a long and robust economic expansion. Nationally, unemployment rates are the lowest they have been in three decades. Economic growth and job creation are robust. Thus, even without any changes in the welfare program, we would expect better economic conditions to draw people into the labor market and off of welfare.

California's recession was deeper and bottomed out later than that of the nation as a whole. Thus, we would expect the national caseload to peak shortly after the economy hit bottom.⁶ Consistent with its deeper and longer recession, we would expect California's caseload to peak slightly later. The caseloads shown in Figures 3.2 (national and California) and the unemployment rates in Figure 3.3 (again, national and California) are consistent with that story.

A similar pattern is apparent across the regions of California. The recession was deeper in Southern California and shallowest in the Bay Area. Figures 3.4 and 3.5 show that the caseload and unemployment rates increased most in Southern California and least in the Bay Area.

⁵ Note, however, that the pre-reform period was not a period of unchanged welfare regulations. There were many changes in the details of welfare programs over this period. See Zellman et al. (1999) for a brief review of the "waiver" period reforms in California.

⁶ Klerman and Haider (1999) discuss why caseload patterns trail economic patterns. In short, flows on to and off of aid are approximately coincident with economic changes. However, the stock of people on aid (i.e., the caseload) adjusts with a lag; people who come on to aid with worsening economic conditions do not all leave immediately.

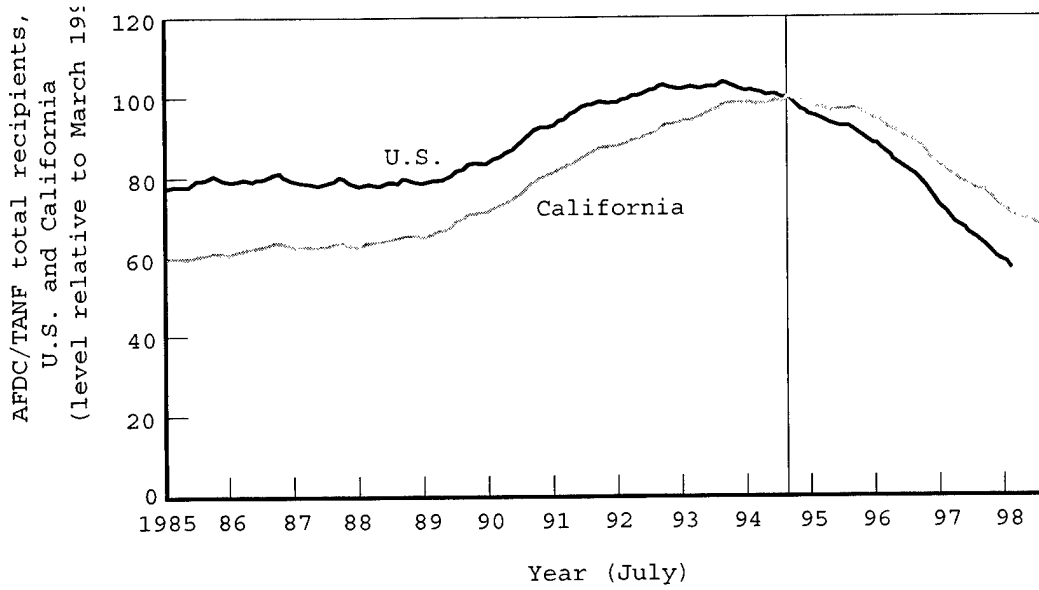


Figure 3.2--National and California Caseloads

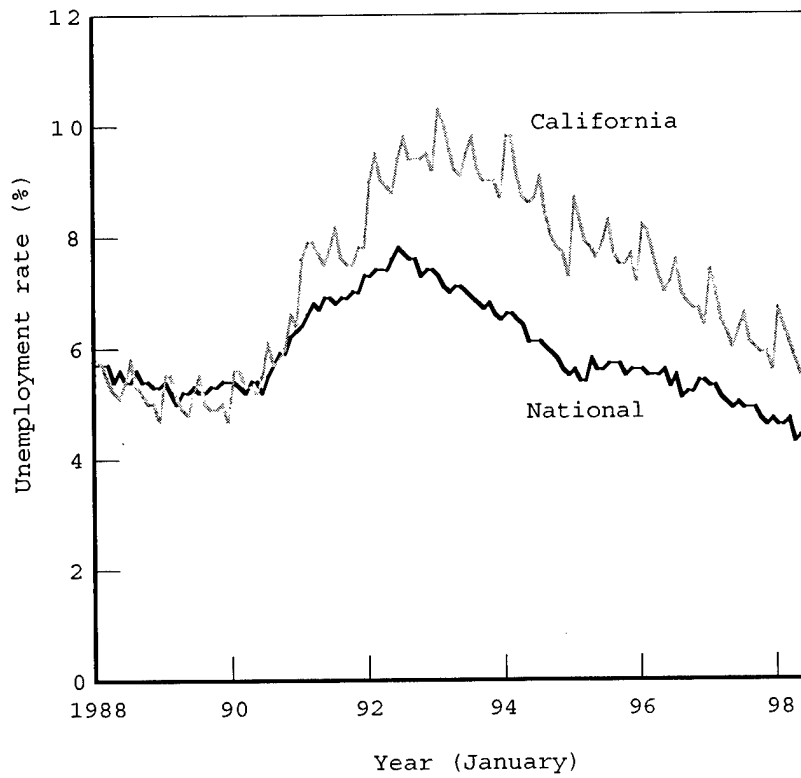


Figure 3.3--National and California Unemployment Rates

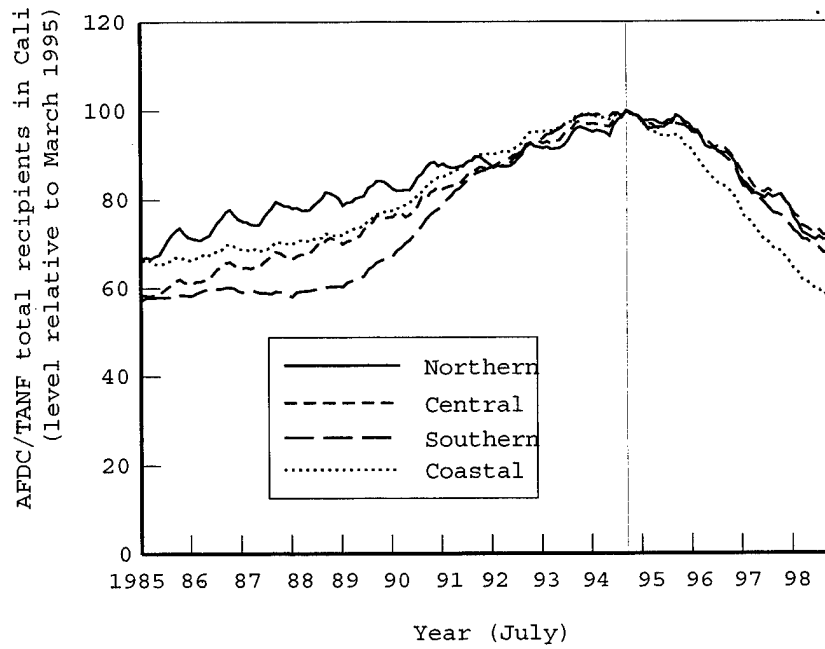


Figure 3.4--California Regional Caseloads

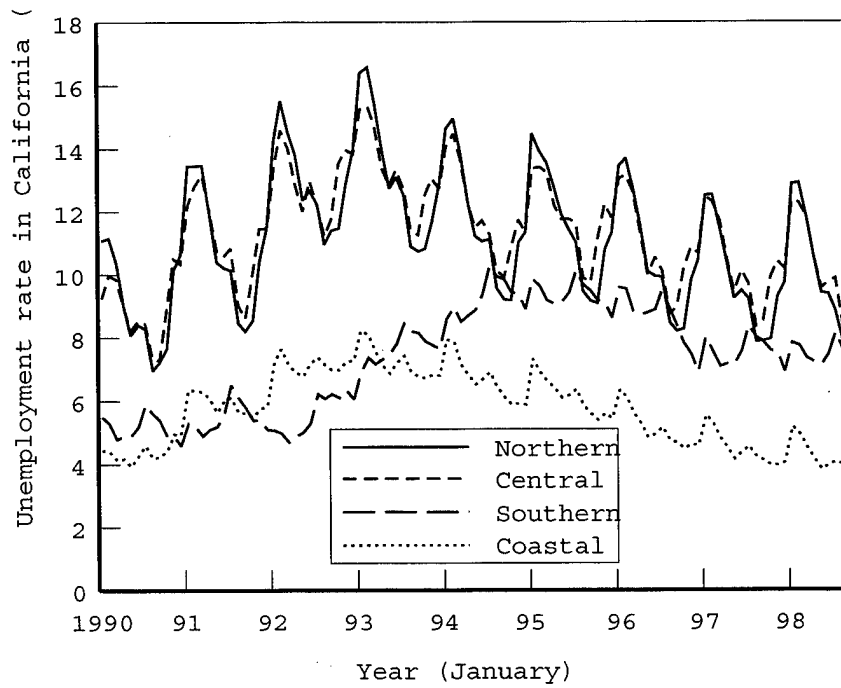


Figure 3.5--California Regional Unemployment Rates

The combination of a priori theory, these simple plots, the national literature (CEA, 1997; 1999; Ziliak et al., 1997; Blank, 1997; Martini and Wiseman, 1997; Wallace and Blank, 1999; Figlio and Ziliak, 1999; Moffitt, 1999), and analyses of California data (Hoynes, 1996; Klerman and Haider, 1999) all suggest that the caseload declines when the economy improves. Therefore, assuming all of the decline since the implementation of CalWORKs as the true CalWORKs effect would overestimate the true effect. The crucial research question is, by how much?

Similarly, what other things are changing at about the same time as the CalWORKs legislation? To isolate the pure effect of the CalWORKs legislation (or of the CalWORKs programs of a given county), we need to back-out the effects of the other things that vary across time and place, factors often referred to as confounders. This is the methodological challenge.

RANDOMIZATION AND CONFOUNDERS

Although defining the general types of net or differential program effects and their specific versions relevant for CalWORKs is relatively straightforward--the three counterfactuals or baselines discussed in Section 2--devising strategies to estimate them is not. This problem of *what would have been* represents the fundamental problem of program evaluation or, more generally, of causal inference.

One way of obtaining unbiased estimates of the counterfactual outcomes in program evaluation is to use an experimental design in which assignment to treatment program or the control program is done at random. That is, one group is randomly assigned to the new program (the "treatment" group) while another group is randomly assigned to the old, or baseline, program (the "control" group). Then, outcomes for the two groups are compared. As a result of the randomization, any difference between outcomes for the two groups must result either from the program (compared to the baseline) or from chance. As long as the sample is large enough, the effect of chance will be small: randomization will eliminate all systematic differences between the treatment and control

groups. Thus, any remaining effect can reasonably be attributed to the differential impact of the two programs.

In the absence of randomization, the situation is much more complicated. Differences in outcomes across two programs could result from the programs' themselves, chance, or because the two groups are very different from each other. When comparing participants and nonparticipants at a given site, several considerations suggest that participants and nonparticipants will be different even before the program begins. Some programs (e.g., the Department of Labor Welfare-to-Work programs) have rules requiring sites to take only the hard to serve. Contractors with performance-based contracts have an incentive to "cream," that is, to take only the easiest to serve to improve their recorded performance measures and thus their income. In voluntary programs, often only the most motivated (and thus most likely to succeed) eligible individuals participate in the program. For some remedial programs (e.g., literacy programs), the less skilled clients will self-select into the program.

These considerations suggest that in the absence of random assignment, simple comparisons of participants and nonparticipants will not yield proper estimates of the true effect of the program. Observed differences across participants and nonparticipants will result from the true effect of the program, chance, and pre-existing differences between participants and nonparticipants. Large enough samples will eliminate the effect of chance. When used, randomization will guarantee that there are no pre-existing systematic differences on average between participants and nonparticipants. When randomization is not used, such differences (confounders) are likely. Thus, estimating the pure effect of the program requires controlling for these pre-existing difference (i.e., controlling for confounders).

When comparing outcomes across geographical areas or through time, related concerns imply similar methodological problems. Even if we placed the identical program in two different places (or different times in the same place), we would expect to find different outcomes. For example, California's counties vary greatly. Some have a more educated work force. Some are more ethnically diverse, or have many refugees.

Some have a rich infrastructure of education and training programs, and some do not. Some have urban mixed economies, while others have rural economies based on agriculture. Some have robust economies with low unemployment rates; others have weak economies with high unemployment rates where few jobs are being created.

Again, we would not expect participants in the same program but in 2 different counties to have the same outcomes (e.g., employment and earnings). This implies that to determine the relative impact of two county welfare programs (or two state welfare programs) *holding everything else constant*, we cannot simply compare outcomes across the two counties (or across the two states). We need to control for pre-existing difference across the counties (or states).

THE NEED FOR NONEXPERIMENTAL APPROACHES TO ESTIMATING CAUSAL EFFECTS

Unfortunately, while randomization is clearly the preferred choice for dealing with the challenge of confounders, it is not an option for evaluating most of the effects of CalWORKs noted in Section 2. Evaluations based on random assignment require that randomization be done as the program is implemented. However, randomization was not built into the early implementation of CalWORKs.

California has used randomization successfully in the evaluation of the GAIN program, Work Pays, and CalLearn. Two considerations led California to not specify a random assignment design for CalWORKs. First, randomization approaches have trouble capturing effects on social norms and general equilibrium effects. CalWORKs is trying to change the expectations of potential recipients with respect to the welfare system and their life choices. Under randomization, people might expect to be assigned to the old program and thus not change their behavior. Similar general equilibrium arguments have been made about entry and deterrent effects, displacement of other trainees, and effects on market wages.⁷

⁷ See the similar discussion in Friedlander et al. (1997, pp. 1819-1823, 1845-1846) and the citations therein. See especially footnote 23 which notes that "[T]hese issues were considered so important that a deliberate decision was made against using a random assignment evaluation design that would create a no-program control group and would therefore interfere with site-wide program coverage."

Second, CalWORKs represents a dramatic restructuring of the welfare system, one that affects not only recipients but also caseworkers, other government employees, and various service providers. Randomization would require that the control group continue to receive the same services from the same system that existed prior to CalWORKs. However, that program no longer exists, and it is far from clear that counties could have kept a scaled-back version of that program in place for long enough to allow a random assignment intervention.⁸

To estimate the causal effect of CalWORKs (or the CalWORKs program in a particular county), then, the Statewide CalWORKs Evaluation must use alternative, nonexperimental approaches to estimate what outcomes would have been under some other program. These counterfactual outcomes can then be compared to observed outcomes to estimate the causal effect of the program. In this section, we consider three such approaches: (1) simple mean difference estimator; (2) regression methods; and (3) statistical matching. As we show below, the first approach is not suitable here, but some combination of the other two approaches shows promise.

Simple Mean Difference Estimator Approach

How would a simple mean difference estimator approach work? For the sake of concreteness, suppose that one wished to estimate the differential effects of the CalWORKs program in one county, County A, relative to that of another county, County B, on the average earnings of welfare recipients. (Accordingly, Treatment 1 ($T_i = 1$) corresponds to the County A program and Treatment 0 ($T_i = 0$) corresponds to the County B program.) To estimate this effect, one could consider using the difference in the average earnings of recipients across these two counties. It is informative to consider in more detail three different reasons why this estimator is not likely to work.

The first reason is the potential noncomparability across the two locations or "environments," over and above differences in the two programs. It is possible, for example, that labor-market conditions

⁸ However, randomization could be used to evaluate particular components of CalWORKs programs.

differ in the two counties. Similarly, differences may exist in social programs or public policies other than welfare. We refer to this as the *environmental heterogeneity problem*. To the extent that these environmental factors are correlated with the likelihood of individuals being CalWORKs recipients and are correlated with the outcomes of interest, $Y(0)$ and $Y(1)$, failure to control for their influence implies that the simple mean difference comparisons suggested above will produce biased estimates of the differential effects of the CalWORKs programs in the two counties.

A second aspect of this problem is the potential for noncomparability in the sets of participants, or subpopulations, enrolled in the two programs. For example, it may be the case that one county's population of welfare recipients is more skilled than the other or has more barriers to work (e.g., physical or mental impairments). Put another way, the distribution of the personal/household characteristics, X_i , varies between the two counties. We refer to this as the *individual heterogeneity problem*. To the extent to which these population differences are correlated with the likelihood of individuals enrolling in a county's CalWORKs program, the simple mean difference estimator again will tend to be a biased estimate of the differential effects of the two programs. In this case, the bias results because the average outcomes would differ across the two counties as a result of the differences in populations, even if there were no true differential effect across the two programs.

A third problem can arise when one wishes to isolate the effects of particular program components. Suppose, for example, that two counties differ in their WTW programs, in that County A assigns all its nonexempt recipients to a "Supported Work" program while County B assigns its nonexempt recipients to a "Human-Capital-Building" program. If these components were the only source of difference in the programs of the two counties (and abstracting from the environmental heterogeneity and nonoverlap problems just noted), then the simple mean differences of outcomes of recipients between the two counties would provide an unbiased estimate of the differential impacts of these two alternative WTW programs. However, the county programs can also differ in other

components, rules, and procedures. In fact, this is likely to be the case for the county-specific CalWORKs programs, given the discretion CalWORKs allows counties in designing their programs. We refer to these differences in other program components as *program component heterogeneity*. In this situation, the simple mean difference estimator is biased for the differential effects of these two programs.

Conventional Linear Regression Approach

Regression analysis is one of many ways to attempt to control for confounders. It relies on strong linearity and additivity assumptions. It works relatively well when three conditions are satisfied. First, the key differences--environmental heterogeneity, individual heterogeneity, and program component heterogeneity--are measured. Second, there are large number of observations of the key differences; for environmental heterogeneity, large numbers of counties; for individual heterogeneity, large numbers of individuals; for program component heterogeneity, again large numbers of counties. Third, the two populations are close in their covariate distributions.

We begin with a conventional regression approach to comparing county programs. Suppose we want to know the effect of a program component that some counties have adopted and others have not (e.g., outsourcing job club or merging eligibility and WTW workers). Furthermore, suppose we have ideal data, recorded outcomes on multiple (large numbers of) counties for (many periods) before and (many periods) after the policy change, for large numbers of people. With less than ideal data, the methodological problems become even more difficult.

An outcome of interest (e.g., earnings, or hours worked) will be regressed on a dummy variable indicating one of the counties and a set of covariates describing other observable differences between the two counties (e.g., the unemployment rate or the characteristics of the caseload). When the policy varies across counties and time, the implied linear regression is

$$y_{ctg} = \alpha + X_{ctg}\beta + Z_{ctg}\gamma + \mu_c + \eta_t + \varepsilon_{ctg} ,$$

where c refers to county, t refers to time period, g refers to subgroup; Y_{ctg} is the outcome of interest; X_{ctg} is the policy of interest; Z_{ctg} are other control variables that vary across counties and across time (e.g., demographic characteristics of the caseload or local economic conditions); μ_c is a dummy variable for county, η_t is a dummy variable for time period; and ε_{ctg} is a regression residual.⁹

Our focus is the effect of the policy of interest, β . The estimates of the coefficients on the other control variables (Z_{ctg}) are not of interest in themselves. Rather, the covariates are included to control for differences in average values of these covariates between the two counties.

Recall our discussion of confounders. These other control variables need to control for environmental heterogeneity (among them economic conditions), individual heterogeneity (e.g., the age of participants), and program component heterogeneity (e.g., other program design differences). Ideally, these control variables will "adjust" for all variation from these forms of heterogeneity. Complete control is never possible. In as much as control is incomplete, we need to worry that there remain uncontrolled for differences between the two groups. In that case, even the "adjusted" comparison of outcomes across programs includes not only the true effect of the program (what we want) and the effect of chance (which will be small if the sample is large enough), but also the effect of remaining uncontrolled for differences in environment, in individual characteristics, and in the programs (what we do not want).

Suppose one of the covariates is age. If the average age of the population of interest in the two counties differs and this is not accounted for, some of the cross-county differences in employment--the part actually the result of the age difference--will be wrongly attributed to the effect of welfare reform. This result induces biases in the estimates of the average effect of the program between the two

⁹ The approach of including dummy variables for each time period is referred to in the econometric literature as the "difference-of-differences" estimation method (Meyer, 1996). Note that this approach requires multiple observations per period. Thus, this approach is not appropriate for state-wide time-series analyses of program effects.

counties. In a regression, including measures of age will at least partially "control" for the differences because of age.¹⁰ It will then be more plausible to attribute the remaining differences to welfare policy.

This example also clarifies why some data are preferred to others. If we have before and after data for each county (or state, or individual), then we can use the county as its own control.¹¹ Instead of simply comparing across counties, we can compare the change as the program is implemented in one county with the change as the program is implemented in another county. In a regression context, including a dummy variable for each county is equivalent to comparing each county with itself. Such dummy variables imply that we do not need to measure all the differences across counties. All time-invariant (or in practice slowly changing) differences across counties will be captured by the dummy variables for each county.

Similarly, with multiple counties, we do not need to control explicitly for changes across time. We can include a dummy variable for each period. This dummy variable will control for all common statewide effects (e.g., state legislation).

This dummy variable approach is not feasible for statewide analyses. To estimate the time dummies, we require multiple observations per period with differing timing of the adoption of new policies. Since CalWORKs replaced AFDC/GAIN nearly simultaneously across counties, the required variation is missing. Instead, we must take the weaker approach of including controls for observed factors that vary over time and county and (perhaps polynomial) time trends. This approach makes the strong assumption that nothing unmeasured changed across the roll-out of CalWORKs and, thus, that all the observed change can be attributed to CalWORKs. This assumption is clearly false. We will need to assess how good an approximation it is. Finally, note that

¹⁰ The control will only be complete if the functional form is exactly correct. The matching methods discussed below are more robust to incorrect specification of functional form.

¹¹ Meyer (1995) discusses these dummy variable strategies in depth.

for county-specific changes, we can include time-period effects, yielding more robust estimates.

The Statistical Matching Approach

The statistical matching approach is an alternative, but not unrelated, approach to regression that attempts to adjust for the differences across comparisons using statistical matching techniques to account for the influences of confounding factors. As its name suggests, linear regression imposes strong linearity assumptions. Thirty-year-olds are implicitly assumed to have outcomes half way between 20-year-olds and 40-year-olds. Such linearity assumptions are often problematic.

Statistical matching approaches relax this linearity assumption. Intuitively, they compare only individuals who are alike in covariates. Rather than assuming that 30-year-olds are half way between 20-year-olds and 40-year-olds, with statistical matching, 40-year-olds are compared to 40-year-olds; 30-year-olds are compared to 30-year-olds; and 20-year-olds are compared to 20-year-olds. With large enough samples and overlapping distributions (i.e., there are 20-, 30-, and 40-years-olds in each county), the linearity assumption is unnecessary. Statistical matching approaches relax this assumption.

In practice, we have many more confounders than age. Even with large samples, it quickly becomes difficult to find enough exact matches (e.g., a 30-year-old, black female, with 10 years of education) to precisely estimate the effects. To address this problem, we group observations based on some metric that measures closeness (i.e., similarity among observations). A particularly appealing metric is based on the *propensity score*. The propensity score reduces the set of covariates required for matching to a single variable--the probability of being in one program rather than the other. Instead of matching on the entire set of covariates, one then matches on this single measure and avoids any bias from differences in covariate distributions. This procedure makes the matching approach feasible, even in the presence of many covariates.

As was true for regression analysis, its main requirement is that the covariates available in the data are sufficiently rich that adjusting for them eliminates all confounders. In other words, comparing individuals with identical covariates is assumed to lead to valid comparisons. As we discuss in more detail below, recent work on these approaches by members of the evaluation team and others suggests that these approaches are promising for recovering the true effects of the policies (e.g., Hotz, Imbens, and Mortimer, 1999; Dehejia and Wahba, 1999; Heckman, Ichimura, Smith, and Todd, 1999).

That available covariates are sufficiently rich to eliminate any effect of confounding factors is a strong assumption. It is rarely exactly true. As we discuss in detail below, we will test the extent to which this assumption biases our estimates by applying these approaches to experimental data. For such data, we know the "truth." By applying these methods to data for earlier welfare reforms in California, we will gain an estimate of the success of the methods for a similar population and similar outcomes.

STRATEGIES FOR (PARTIALLY) VALIDATING THE METHODS

As noted above, the use of either regression or matching methods is not guaranteed to eliminate the confounding influences of environmental heterogeneity, individual/household heterogeneity, and program component heterogeneity. Thus, it is essential to gain some sense of how well these methods work and, more importantly, which types of factors must be controlled for in our regression and matching analyses to reduce resulting biases. As we get a sense of when the methods succeed, including how large a sample is required and what effects must be controlled for, we can identify for which outcomes the data will be sufficient to estimate causal effects.

Several recent studies suggest the use of matching methods shows some promise, although their findings are not uniformly positive. In one study, Heckman and Hotz (1989) find that they can use regression-based methods to adjust for differences between the randomly assigned control group in the National Supported Work Demonstration project and comparison groups derived from Current Population Survey (CPS) data to

eliminate the sources of selection bias noted by LaLonde (1986) in his earlier study of this data. The important feature of the Heckman and Hotz study is that they demonstrate that one can use a variety of hypothesis-testing strategies, many of which can be implemented without the benefit of experimental data, to choose appropriate regression methods for alignment of the outcomes for control groups and nonexperimentally generated control groups.

In recent work of Dehejia and Wahba (1999), the propensity-score methodology, originally based on work by Rosenbaum and Rubin (1983), has been applied to employment training programs. Dehejia and Wahba consider the same National Supported Work Demonstration data used by LaLonde and Heckman and Hotz. Using the nonexperimental propensity-score controls, Dehejia and Wahba were able to estimate the same program effects as were estimated through random assignment.

At the same time, a recent study by Heckman, Ichimura, Smith, and Todd (1998) analyzes the use of propensity-score methods to align the outcomes for control groups from the National JTPA Study with comparison group data constructed from data on JTPA-eligible nonparticipants drawn in the various localities in which the study was conducted. They find that propensity scores do not work very well to align outcomes. Furthermore, they find that the reason for this failure is the lack of comparability, or overlap, between the eligible nonparticipant subpopulations and those members of the control group who actually applied for JTPA. This work clearly highlights the importance of analyzing the "overlap" issue in the analyses to be conducted in RAND's evaluation of CalWORKs' effects.

Currently, RAND team members Hotz and Imbens, in conjunction with Julie Mortimer, a graduate student at UCLA, are exploring the validity of this matching strategy for use in the CalWORKs evaluation. In earlier work related to this project, these authors analyzed data from WIN demonstration experiments conducted by MDRC during the 1980s. Using these welfare reform demonstration projects that did use random assignment, Hotz, Imbens, and Mortimer (1999) apply propensity-score and regression methods of the type outlined above to form matched samples for the "treatment" group in one state/region with individuals from

other states/regions. Various sets of characteristics, including past histories of welfare participation and work, were used to construct the matches. To assess the reliability of these matches, the researchers compared the distributions of outcomes for the matched samples with the control and the treatment groups generated by random assignment. They find that by conditioning on past-earnings histories, on a limited set of personal characteristics such as age and gender of the household head, and on a small number of measures of local labor-market conditions, they can align the average outcomes (they analyze earnings and welfare participation as their outcome measures) for the control group in one WIN program (San Diego's SWIM program) with those for control groups for programs in other locations (Arkansas, Baltimore, and Virginia).

This conclusion holds whether they use propensity-score or regression methods. This implies that it may be possible to use matching methods to project what outcomes would have been under the old AFDC/GAIN program. Outcomes under the new CalWORKs program are observed. Thus, we may be able to estimate the effect of CalWORKs relative to GAIN. At the same time, these authors find that they cannot align the outcomes for the treatment groups across programs. They hypothesize that this result is caused by the differences in program treatment components (i.e., program component heterogeneity) across these programs. In the WIN data, they lack information on program components available in the various programs analyzed, so they cannot adjust for such measures.

As part of RAND's evaluation of CalWORKs, Hotz, Imbens, and Mortimer are currently conducting a similar analysis of data on California counties from the MDRC GAIN evaluations. The analysis of the GAIN data is significant for several reasons. First, it contains data on California populations and programs. Second, long-term follow-up data are available. Third, the project may be able to get access to detailed information on GAIN program components. The latter type of data was not available in the data for the WIN evaluations. Findings to date closely parallel the results from their analysis of WIN data. In particular, they again find that matching methods which condition on

work histories and a limited set of personal characteristics allow one to align the GAIN control groups in different counties. At the same time, without controlling for measures of county-specific GAIN program components, they are not able to align the outcomes for experimental groups.

OTHER TECHNICAL ISSUES IN USING THE TWO NONEXPERIMENTAL APPROACHES

Using the two approaches discussed above--regression and statistical matching--presents three technical issues: (1) additional data requirements, (2) stratification, and (3) the form of the statistical model. Each is discussed below.

Additional Data Requirements

Implementing either the regression approach or the matching approach imposes three additional data requirements. First, if we are to estimate the effects of different CalWORKs implementations, we will need to be able to characterize the implementations. The All-County Implementation Survey (ACIS) being conducted as part of RAND's evaluation is collecting some information on the CalWORKs implementation in each county. A survey of caseworkers in 21 counties is collecting more information on implementation. For the six focus counties and eighteen follow-up counties, we will have additional information from key informant interviews. As we identify important variation across the counties, we will add more questions to the ACIS and the process analysis to refine our understanding of the differences.

Second, we need to measure rates, which are the ratio of an outcome to the population at risk. The outcomes can often be estimated from administrative records. We will estimate the size of the population at risk using estimates of the population of the state by county, gender, and age. Such estimates are available from the State Demographer and from the U.S. Bureau of the Census, as well as from private firms. We are currently evaluating the relative merits of each source.

Third, to control for confounders, we need to measure them. Through the ACIS, site visits, and reviews of the secondary literature, we will compile a database of other potentially important differences in county policies (e.g., other state demonstration programs) and

characteristics. Similarly, we will construct refined estimates of local labor market conditions and other county characteristics.

At the individual level, we will want to control for--or stratify by (see below)--individual characteristics. The exact items to be controlled for will vary from analysis to analysis. Among the items to be controlled for are program type (one-parent, two-parent and child-only), demographics (gender, age, race/ethnicity/language, immigrant status, education, literacy, number of children, and age at first birth), history of aid receipt, employment history, and barriers to employment (physical disabilities, mental illness, and substance abuse).

As noted earlier, such control variables are also crucial for identifying the subpopulations in which we expect effects to be largest and the subpopulations that can be used as "control groups." Unfortunately, because some of the administrative data sets are missing even such basic demographic information as gender, age, number of children, and marital status, controlling for confounders and identifying subpopulations is not possible for some outcomes and some sub-populations. As we discuss in Section 4, we are exploring data-matching schemes that would allow us to append some of the missing covariate information for at least some subpopulations, but none currently appear promising.

Stratification

As appropriate and given data availability and sample size, we will stratify our analyses. Such stratification will allow us to consider how outcomes vary across subpopulations. Among the stratifications of interest are type of AFDC/TANF case (FG, UP, and child-only), history of welfare, those who are especially vulnerable, and demographic variables.

We will proceed with caution when stratifying on individual receipt of other services to estimate the "effect" of the services. It is not clear that available controls (e.g., regression and matching) will be sufficient to control for program rules. Sometimes, services will be offered to those viewed as most employable (i.e., "creaming"). In that case, comparisons across those who do and do not receive the services will yield an overly positive estimate of the effect of the services on

outcomes compared to the desired effect of offering or denying the services to a given individual. Sometimes, the services will be offered only to those with the corresponding problem (e.g., alcohol abuse) or to those viewed as least employable (e.g., Department of Labor WTW funds). In this case, comparisons across those who do and do not receive the services will yield an overly negative estimate of the effect of the services on outcomes.

Instead, we will compare the effects on outcomes across counties with different policies concerning the provision of services. Here, our methods for the control of confounders are likely to be more effective. Making this comparison will require detailed data on policies about the provision of services. We will collect this information through the process study and the ACIS. To the extent that we can identify the at-risk population, these analyses will be stronger.

The Form of the Statistical Model

The preceding discussion of confounders and approaches to controlling for them was developed for the simplest case--continuous outcomes (e.g., earnings of current recipients). For that case, standard linear regression and matching strategies are directly applicable. Other types of outcomes are more appropriately analyzed using other statistical models. Binary outcomes are often better analyzed with probit or logit regression models. Intake rates are often better analyzed using grouped data and count data models. Processes occurring in time are often better analyzed using hazard models. Selection and implementation of a statistical model appropriate for the type of outcome is usually straightforward. When we discuss outcomes and data sources in Section 4, we will discuss these other modeling issues.

In contrast, controlling for confounders remains a substantial methodological problem, no matter what the type of outcome. The general approaches discussed above--regression and matching--can be applied to any of these types of outcomes. We will do so, as appropriate.

4. DATA AND OUTCOMES

The previous two sections have discussed what we want to know, the available methods for estimating causal effects, and the general data requirements. In this section, we review the data sources available for each outcome and the implications of the characteristics of the data for analysis. Following an overview of characteristics of the available data sources and a discussion of two data sets that have some information on a broad range of outcomes--the Current Population Survey (CPS) and the Six County Household Survey (6CHS)--we organize our presentation around the data sets for the four types of outcomes: 1) welfare system outcomes; (2) self-sufficiency and employment outcomes; (3) family and child well-being outcomes; (4) financial outcomes.

DATA SET CHARACTERISTICS

We have already noted the close connection between the characteristics of data sets and the possible analyses. Before turning to the outcomes, we briefly review the characteristics of each data set. These data characteristics constrain the analyses we can do.

An ideal data set would have data for both before and after CalWORKs, for every person (not merely a sample), for each of California's 58 counties (or all 50 states), and for current, former, and potential future recipients. Of course, no data set is ideal on each of these criteria.

Table 4.1 summarizes the characteristics of the primary data sources we currently propose to use in our impact analysis. The first row considers whether data are available for both before and after the inception of CalWORKs (B/A) or only for after (A). We note that, with the exceptions of the Six-County Welfare Administrative Data (6CWAD), the 6CHS, and Child Welfare (CWS/CMS) system/case management system, all these primary data sources have data available from both before and after CalWORKs. The second row considers sample size. With the exception of the CPS, Q5, and the 6CHS, all the data sources contain

records for the "universe," rather than simply for a sample.' The third row considers the coverage of the data sources. Most of the data sources are available statewide. Q5 data are available only for the 19 largest counties and then only for about 300 cases per county per year. As their names suggest, two of the data sources--6CWAD and 6CHS--are only available for the six focus counties. The CPS is a special case, which we will discuss in the next subsection. The fourth row considers the set of people covered--current, former, and potential recipients. How this information can be used is highly dependent on the outcome of interest. We defer discussion of these rows until our discussion of the outcomes.

Table 4.1
Characteristics of Primary Data Sources

Data Characteristics	CPS*	6CHS	MEDS	Q5	6CWAD	MEDS-EDD	EDD
Before/After	B/A	A	B/A	B/A	A	B/A	B/A
Sample/Universe	S	S	U	S	U	U	U
Counties Covered	*	6	58	19	6	58	58
Type of Recipients	C,F,P	C,F	C	C	C	C,F	C,F,P

Abbreviations: B/A = before and after CalWORKs; A = only after CalWORKs; S = sample; U = universe; C = current recipient; F = former recipient; and P = potential recipient; CPS = Current Population Survey; 6CHS = Six County Household Survey; MEDS = Medi-Cal Eligibility Determination System; Q5 = Quality Control data; 6CWAD = Six County Welfare Administrative Data systems; MEDS-EDD = MEDS-Employment Development Department earnings match; EDD = Employment Development Department earnings match.

Notes: * CPS could cover all counties, but does not cover all counties in each year, and only the largest MSAs are identified (counties are never identified).

This table lists only the primary data sources. We consider them "primary" because they are the most readily available and because they have the most advantageous characteristics. We have explored and will continue to explore many other data sets. However, as detailed below, our preliminary review suggests that other data sources are less attractive along these dimensions than the ones chosen. For example, other data sets often contain only post-reform data (or the data are not

consistent from pre-reform to post-reform) or they may have only small samples. In addition, they often cover only a small number of counties (sometimes, just a single county), do not cover even most of the focus counties, do not identify counties at all, or are often limited about which populations within a county they cover.

For some of the outcomes of interest (e.g., crime, school attendance, and graduation rates), aggregate data are available. Unfortunately, such aggregate data do not separately identify CalWORKs recipients. Given the other changes in the state--especially the robust economy--such aggregate information alone is not enough to describe post-CalWORKs outcomes for a relevant subpopulation (e.g., current recipients or recent recipients) and is certainly not sufficient to estimate the causal effect of CalWORKs.

THE CPS: THE PRIMARY DATA SET FOR INTERSTATE COMPARISONS

Large national, general-purpose surveys collect considerable information on California and on the welfare population, although this is not the primary focus of these surveys. California has slightly more than 10 percent of the nation's population. In California, slightly less than 10 percent of the population receives cash assistance. Thus, approximately 1 percent of a national random sample would be expected to be current welfare recipients in California. In practice, public assistance often appears to be underreported in surveys.

We propose to analyze the largest ongoing national, general-purpose survey--the U.S. Bureau of the Census's Demographic Supplement to the March CPS. The Census Bureau surveys approximately 50,000 households each March, so we would expect more than 5,000 California families, 500 current welfare recipients, and several hundred more recent welfare recipients. In addition, because of the CPS's rotation group structure, some analyses can be done on a sample three times as large. This is still only a sample of moderate size: Within the state, only the largest metropolitan areas are identified, and no counties are identified.

Despite these moderate sample sizes, the CPS has several important advantages. First, it contains detailed information on some important

outcomes. It contains some information on program participation, such as receipt in the previous calendar year of TANF, Supplemental Security Income, Food Stamps, and Medicaid/Medi-Cal. It also contains detailed information on employment in the last week--such as actual hours, usual hours, type of employer, hourly wages, and earnings--and employment, hours, and earnings in the last year. Finally, it contains some information about child and family well-being. That information includes detailed information on health insurance coverage, family structure and marital status, and income relative to the poverty line.

Second, the CPS has been operating in nearly its present form for several decades. Thus, considerable pre-reform data exist. Furthermore, the data are well understood, relatively easy to work with, and released only several months after collection.

Third, the CPS is a national survey. Thus, we can use the data to conduct interstate descriptive analyses and to estimate causal effects, using other states as the baseline. Because of its national coverage, the CPS will be used by many other analyses across the nation for national analyses. Using the CPS data, we will be able to reexamine those national analyses from a California perspective (e.g., to determine the implied outcome in California, given the outcomes in other states).

Fourth, and perhaps most important, the CPS is a general population survey. As such, it contains information not only on current welfare recipients but also on potential future recipients. We argued earlier that future recipients are a crucial group for which to explore effects, and also a group we otherwise have difficulty measuring. Our ability to distinguish current recipients from recent (in the previous calendar year) is, however, limited.

Additional national survey data and detailed administrative data on outcomes of interest might be available. The Survey of Income and Program Participation (SIPP) and the Survey of Program Dynamics (SPD) contain more information on program participation and on child and family outcomes.¹² The Health Interview Survey contains more

¹² The SIPP is an alternative source of national data. MaCurdy and O'Brien-Strain (1997) analyzed the data to project the effects of

information on health status. National birth certificate data would have more information on fertility, marital context of fertility, and early child health.

However, our tentative plans do not involve analyzing any national data beyond the aggregate reporting to the U.S. Department of Health and Human Services (discussed below) and the CPS. This is true for several reasons. First, the data lag is such that by the end of the evaluation, only early post-TANF data would be available. Second, several highly qualified national research teams are exploring these data. Third, processing each additional data source is expensive. In summary, while we will survey the developing national literature, extensive analysis of national data does not seem to be the best use of contract resources.

THE 6CHS: THE PRIMARY DATA SET FOR CHILD AND FAMILY OUTCOMES

The major primary data available for conducting the impact analysis are state and county welfare administrative data systems and information on earnings from unemployment insurance and tax filings. However, these administrative data are insufficient to address all the outcomes of interest. In particular, as we will show below, information on child and family outcomes for current recipients is very poor and is even worse for former or potential recipients.

To fill in these deficiencies, RAND is fielding the 6CHS effort within the six focus counties specified by CDSS: Alameda, Butte,

welfare reform in California. Also conducted by the U.S. Bureau of the Census, the SIPP is of approximately the same size and has more detailed information on some outcomes of interest. In particular, it follows respondents longitudinally, making it possible to track changes in individual outcomes through time. However, it has several major disadvantages. First, and most important, the delay between data collection and data release is much longer than it is for the CPS. Other disadvantages include problems of sample attrition and a very complicated data structure.

The SPD is another possibility. Also collected by the U.S. Bureau of the Census, the SPD was specifically funded by Congress as part of PRWORA to better understand the effects of reform. Unfortunately, the data-collection effort has several important flaws. First, there are serious problems of differential attrition. Second, as a result of question wording changes, the data are not consistent across periods. Some of the data are collected prospectively, while some are collected retrospectively. Finally, the data have been released only slowly.

Fresno, Los Angeles, Sacramento, and San Diego. The 6CHS will interview approximately 475 current and recent recipients in each of the six focus counties (under current assumptions about response rates). We defer until the next section a more detailed description of the 6CHS and related issues.

WELFARE SYSTEM OUTCOMES

The immediate outcomes of interest in the CalWORKs reforms are in the welfare system and include caseloads, costs, program activities, compliance, and sanctions. In this subsection, we discuss these outcomes and the available data.

Table 4.2 summarizes the available data. It shows that we have information on receipt of cash aid and Medi-Cal from several sources, including the nearly ideal MEDS data. Information on Medi-Cal and participation in CalWORKs activities and receipt of CalWORKs services is available only in the Q5, 6CWAD and in the 6CHS.

Table 4.2

Use of Various Data Sources—Welfare System Outcomes

Specific Outcomes	CPS	6CHS	MEDS	Q5	6CWAD	MEDS-EDD
Caseloads	X	X	X	X	X	X
Caseload Dynamics			X	X	X	X
Aid Payments				X	X	
Program Activities				X	X	*

Notes: X = The data contain this element (subject to quality assessment); * = Earnings indicate work as a program activity.

Caseloads

Caseloads are the most tracked and most widely reported welfare-system outcomes. Counties report caseload data, disaggregated by FG and UP, monthly on the CA237 form. The data are available with an approximately a four-month lag. We have acquired and processed the detailed data back to 1992 and have procedures in place to receive updates monthly.

California and the other states are required to report some outcomes to the U.S. Department of Health and Human Services; these

include caseloads (and some characteristics), total aid payments, total expenditures, and participation rates. These limited data are readily available, as is some basic information on TANF program components. We will perform some simple analyses of causal effects using such data.

Complementing the aggregate data from the CA237 form is the MEDS, which, as its name implies, exists primarily to verify eligibility for Medi-Cal. As such, it contains an eligibility code, which identifies AFDC/CalWORKs and FG/UP. The file also includes basic demographic information--gender, age, race, ethnicity and, recently, language. There is no information on CalWORKs activities, the receipt of services, or the amount of the aid payment.

MEDS data are available monthly, with one record per eligible individual (adults and children). New cases are reported with a lag of several months, so the most recent months are not usable for purposes requiring the most up-to-date information. We have constructed a simple forecasting model that allows us to extrapolate (or "complete") the final caseload and its characteristics for each month from the preliminary and incomplete early MEDS data.

We have acquired and processed historical data for January of each calendar year back to 1987. The lagged reporting problem implies that June data are needed to fill in data for those on aid only for a short time or for those who started late in the calendar year. We are now in the process of acquiring and processing the data.

Using these MEDS data, we can describe caseload trends overall and disaggregated by program type, demography, and dynamics of the caseload. A companion volume, Haider et al., (1999), provides a more complete description of the CA237 and MEDS data, discusses the completion and forecasting model, provides some early disaggregated tabulations, and reports some prototype dynamic analyses.

From an analytic perspective, the MEDS data are nearly ideal for analyzing the number of cases. We have consistent pre- and post-CalWORKs data for the entire population (not a sample) and program (FG versus UP versus child-only), and the county of residence is identified. Thus, we can consider both AFDC and inter-county baselines using both regression and matching approaches. Furthermore, the data identify the

county of residence and basic demographics. Thus, we can do the analyses disaggregated by demographic group and program. Given the importance of the outcome, we plan to do so using the MEDS data, using the CA237 data primarily as a data check.

The CA237 data and the aggregated MEDS data are counts. While this might suggest using a Poisson regression model, the counts are large enough that simply modeling rates--cases per person in the population using linear regression with a weighting correction for cell size--will be more than sufficient. Given the availability of covariates in the MEDS, these analyses can be disaggregated. Haider et al. (1999) contains a preliminary analysis of the most salient dimensions for disaggregation. Two approaches are possible--complete stratification or regression modeling with multiple high-order interactions. We expect to explore both approaches.

Creating the rates requires estimates of the population at risk. For disaggregated or stratified analyses, we require population estimates stratified in ways consistent with the disaggregation in the MEDS. Population estimates are available from three sources: the state demographer, the U.S. Bureau of the Census, and private forecasting groups. These estimates include some combination of county of residence, age (often grouped), gender, and race-ethnicity. We are currently exploring the relative advantages of the various data sources.

Caseload Dynamics

Beyond the number of cases, we also want to know their dynamic character and how the cases are evolving. For example, we want to know how the level of new awards is changing, how the share of the caseload at different durations (under three months, four to twelve months, one to three years, three to five years, and five or more years) is changing, how the hazard rates are changing (i.e., the probability that a one-month-old case, a 12-month-old case, and a 60-month-old case will leave the rolls in the next month), and how quickly recipients are accumulating time against the TANF 60-month time limit.

Two sources of data on such caseload dynamics are available. First, the CA237 form (discussed above) reports applications and new

cases. From these two data elements, we can compute the share of applications approved. This measure gives a rough proxy for the strength of screening at intake and the percent of applications approved. The measure is not a perfect measure for the strength of screening. Once word "hits the street," application behavior itself may react (or even overreact) to the approval process. Inasmuch as such self-selection occurs, the approval rate will be much less informative. The process analysis will need to be attentive to such claims.

Second, by linking the MEDS files together by social security number (SSN), we can construct lifetime histories of aid receipt for individuals back to 1987.¹³ We can therefore estimate the probability of first entering aid, the probability of leaving aid conditional on time on aid (i.e., the hazard rate for exit), and the reentry rate conditional on time off of aid (i.e., the hazard rate for reentry). Each of these transition probabilities can be allowed to vary by individual characteristics, county, county program, and county economic conditions. These estimates can then be combined to yield estimates of the caseload at a point in time and in the steady-state. Given the available data and the importance of dynamic characteristics, it is possible to thoroughly describe and estimate causal effects against multiple baselines using multiple methods, and we plan to do so using the MEDS data. Again, the CA237 data will be used primarily as a data check.

The appropriate methods for these dynamic analyses vary. Applications, approvals, and the size of the caseload by duration can be analyzed by the rate-based regression approaches discussed with respect to the caseload. To analyze the hazard rates, we will use discrete time-hazard models. From the results of the estimation, we will construct steady-state caseloads.

¹³ Constructing case histories (in addition to individual histories) requires stronger assumptions and does not appear to be as useful.

Aid Payments

From a financial perspective, the level of aid payments is at least as important as the number of cases. Data on aid payments, however, are more limited than data on caseloads. Monthly in the CA237, counties report total aid payments separately for FG and UP cases. Combined with caseload data, this information allows the estimation of payment per case. For these aggregate data, we will describe the outcomes and estimate the effects of CalWORKs relative to AFDC/GAIN and across counties.

Beyond simple aggregate calculations, sufficient statewide data are not available to do detailed individual-level analyses of aid payments. In particular, statewide individual-level data on payments per case are not available. Detailed information on the aid payment and how it was computed has traditionally been a key component of the quality assurance systems. Q5 and its predecessors provide information for about 300 cases per year for each of the California's 19 largest counties. An additional 300 cases are spread throughout California's 39 smaller counties. Although we plan to explore the Q5 data and its predecessors, preliminary indications are not encouraging. First, the sample sizes are small. Second, as noted, coverage of counties is incomplete. Third, and most important, state officials have expressed serious concern about the quality of the data in the early CalWORKs period. These concerns appear to be least salient with respect to the payment data.

In the six focus counties, we expect to measure individual benefit payments from the county administrative data. We are still exploring those data, so it is too early to make precise statements about the analyses that can be done. Our preliminary investigations suggest we will have individual-level data on the size of the aid payment and the main factors (size of the assistance unit, exempt/non-exempt status, labor earnings, other income, sanction status) determining that payment from early 1998. The availability of earlier data is less clear. The lack of pre-CalWORKs data implies we will use the county data primarily for description and to better understand the aggregate data. Full

estimates of the effect of CalWORKs on disaggregated aid payments seem unlikely.

Program Activities

Beyond simply measuring the number of cases, their characteristics, and the size of aid payments, we want to know what actually happens to CalWORKs participants while they receive cash assistance. In particular, we want to know in which program activities they participate; what program activities they were summoned to but did not participate in, and whether they were sanctioned; how they are meeting their work activity requirements (if at all); and what support services they are receiving.¹⁴

Again, information on these outcomes is available from county reporting requirements and from county administrative data. Counties were required to report participation in their GAIN/WTW programs on the GAIN25 form. We have begun acquiring and analyzing these data. We do so cognizant of the fact that state officials have expressed considerable concern about the quality of these data.

Furthermore, the GAIN25 data for 1999 appear to be problematic. Shortly after the CalWORKs legislation passed, a working group developed a WTW25 form to update the GAIN25 form to reflect the program changes. After several false starts and more than a year, the GAIN25 form was to be replaced by the WTW25 form in July 1999. (See ACL-99-24, April 24, 1999.) With the establishment of the Separate State Program (SSP) for two-parent families, that form was to be changed yet again. (See ACL-99-60, September 2, 1999.) Preliminary indications are that counties are having trouble completing the form and there is likely to be a period of several months in which no data are available for some counties and the data for other counties are of very poor quality. Unfortunately, the last months of the GAIN25 data and the early months of the WTW25 data cover crucial months in the development of the CalWORKs program.

¹⁴ See Zellman et al., 1999, on the importance of noncompliance and sanctions.

Two other sources of official reporting data are also available. First, in addition to reporting the GAIN25 data, counties in the CalWORKs period are required to report their participation rates. Since reporting has just begun, it is not clear how useful these data will be. Second, Q5 collects extensive data on program activities. The earlier noted concerns about the Q5 data are most salient for these outcomes.

For the six focus counties, we are currently acquiring the county GAIN data systems. Again, preliminary indications are that data will be available back to early 1998, but perhaps not earlier. We anticipate that considerable information will be available for each county. However, exactly which data will be available are less clear; however, it seems likely that the data will not be totally comparable across the counties, nor consistent through time.

Current indications are that actually processing this data will be very expensive. Understanding how the data are coded and what data are reliable is likely to require extensive interaction with county coordinators and other county staff. Understanding what goes on inside the "black box" is crucial to the process analysis, the impact analysis, and the cost-benefit analysis. Thus, this data-processing task has first claim on data-processing resources. Until we understand just how expensive it will be, we are reluctant to analyze a large number of other data sets of less clear importance.

SELF-SUFFICIENCY OUTCOMES

By requiring work, PRWORA and CalWORKs embody a clear model for decreasing the caseload and aid payments. Through work, almost all participants are expected to achieve employment and high enough wages to stop receiving cash aid before reaching the five-year lifetime limit on adult cash aid receipt. CalWORKs provides extensive WTW activities, case management, and support services to help participants achieve this goal. Still, given the skills of current recipients, the outcomes of even the most successful JOBS/GAIN programs suggest that achieving self-sufficiency before time limits will be a major challenge.¹⁵ The CalWORKs evaluation should provide both measurements of the level of

¹⁵ See the discussion in Zellman et al., 1999, pp. 52-56.

employment and earnings and estimates of the effects of CalWORKs programs on those outcomes.

Table 4.3 summarizes the available data for analysis of self-sufficiency outcomes. Nearly ideal data are available on employment and earnings for current and former recipients from the MEDS-EDD match. Some data are available on hours of work and hourly wage from the CPS, Q5, 6CWAD, and the 6CHS.

Table 4.3
Use Of Various Data Sources--Self-Sufficiency Outcomes

Specific Outcomes	CPS	6CHS	MEDS	Q5	6CWAD	MEDS-EDD
Employment and Earnings of Current Recipients	X	X			X	X
Employment and Earnings of Past Recipients		X				X
Employment and Earnings of Potential Future Recipients	X					
Hours of Work and Wages	X	X		X	X	

Notes: X = The data contain this element (subject to quality assessment)

Employment and Earnings of Current and Past Recipients

CDSS's MEDS-EDD provides high-quality data on employment and earnings for all current and past recipients of cash aid (AFDC/CalWORKs) (age 16 and over). The data are compiled from the unemployment insurance filings of individual firms, report employment, earnings, and an employer identification number for all covered employment for each calendar quarter. Covered employment includes about 90 percent of all jobs. Only federal government employees, the self-employed, and "under the table" employment are not included.¹⁶ The data can be combined across quarters to estimate employer tenure and earnings growth (through time, across employers, and within employers). Data are readily available back to 1992, and we have begun to process the data. Some pre-1992 data are available, but it appears that 1990 and 1991 data are

¹⁶ Comparison with grouped FTB tax return data and matched 6CHS data should allow some evaluation of the completeness of the EDD data.

not recoverable. Therefore, we expect to analyze data from 1992 forward.

Our current plans involve extensive analysis of these data. The sample sizes are large; the data cover all individuals who have received aid since 1992; the data cover each of the state's 58 counties; and historical data are available back well before CalWORKs. Therefore, we expect to be able to both describe outcomes and estimate causal effects both relative to AFDC/GAIN and across counties.

The data are rich enough to allow the analysis of several important outcomes.¹⁷ We will begin with whether the participant is employed at all. We will then explore average quarterly earnings and the proportion of individuals with earnings greater than cutoff values, e.g., half-time employment at the minimum wage, full-time employment at the minimum wage, and self-sufficiency (full-time employment at a wage high enough to be ineligible for cash assistance).

Since the EDD earnings data are matched to the MEDS, we can explore how employment and earnings vary across subpopulations. We can do these analyses stratifying by demographics (e.g., race-ethnicity, number of children). We can also consider how these outcomes vary dynamically. For example, we can consider how employment and earnings vary with time since first receipt of aid (since 1992), since the beginning of the most recent spell of aid receipt, since the last time an individual left aid, and since a particular departure from aid (even if there is a subsequent return) and we can consider whether earnings are growing through time, across employers, and within employers. Appropriate methods are straightforward linear regression and binary outcome regression (e.g., probit, logit, linear probability model).

Employment and Earnings of Potential Future Recipients

One pathway through which PRWORA and CalWORKs might affect outcomes is by discouraging individuals who might otherwise have gone on aid from

¹⁷ These MEDS-EDD data, however, do not include information on hours of work. For reasons we discuss below (at "Household Resources and Poverty"), MEDS-EDD data also do not include information on total resources available to the household in which the children live (e.g., spouse's earnings).

ever receiving cash aid. To measure such effects, we would ideally track employment and earnings of potential future aid recipients. Such potential future recipients are likely to be concentrated among young females.¹⁸

Obtaining information on the employment and earnings of potential future recipients is not as straightforward as is obtaining such information for current and past recipients. The problem is not EDD data. In principle, EDD data are available for anyone working at a covered job in California. Gaining access to the EDD data for such a comparison group would require that EDD create a new extract from its data. Some additional negotiations would be necessary, but we are optimistic that they could be successfully concluded. Instead, the problem is that, by themselves, the EDD data include only an SSN, an employer ID, and the amount of earnings. There are no demographics. Thus, we cannot identify young women, and certainly not women who have recently given birth (at all or a first birth).

The EDD data include SSNs. Therefore, if we could identify and obtain access to a "donor file" with SSNs, we could track the employment and earnings of potential future recipients. We have identified four candidate donor files: (1) Department of Motor Vehicles drivers license data,¹⁹ (2) Social Security Administration application data,²⁰ (3) birth certificate data, and (4) non-CalWORKs Medi-Cal adults. Each of these approaches has problems.²¹ After receiving the comments of the

¹⁸ Some have argued that effects should be concentrated among women recently giving birth. While we agree with this perspective, we note that PRWORA explicitly aims to discourage those births, so women recently giving birth may be too narrow a set of potential future recipients and thus may underestimate the effects of CalWORKs.

¹⁹ Department of Motor Vehicles driver's license data include gender, age, and race-ethnicity. We have begun the process of obtaining permission to do the required match. We are, however, concerned about differential possession of driver's licenses, especially in poor populations.

²⁰ The Social Security Administration has information on gender, age, and race, for essentially the entire population from applications for SSNs. We have begun exploring the possibility of access to such data. Preliminary indications are not promising, but considerable work remains to be done.

²¹ There is some prospect of matching directly to birth certificate files to identify recent births and first births. It appears that birth

Technical Subcommittee, we have given these efforts a lower priority. We do plan some descriptive analyses using the CPS.

Hours of Work and Wages

Although the EDD data provide information on employment and earnings, they contain no information on hours of work or hourly wages. Hours of work, however, are a crucial component of federal and state participation requirements, and hourly wages are a standard measure of earnings potential.

Information on hours of work and wages should be available from several sources: the 6CHS for current and recent recipients; the Q5 (in 19 counties) and the administrative data (in the six focus counties) for current recipients; and the CPS statewide for current, recent, and potential future recipients. None of these data sources are ideal. They suffer from some combination of small samples, incomplete coverage (e.g., not recent or potential future recipients), and non-identification of counties. Together, this leads us to conclude that we will perform some descriptive tabulations, but the data will not support a thorough description of wages or any serious analysis of causal effects of the legislation on hours or wages.

CHILD AND FAMILY WELL-BEING OUTCOMES

An assessment of CalWORKs will need to consider not merely welfare system and self-sufficiency outcomes but also its effects on children and their families. Unfortunately, the data available for such an assessment are weaker than those available in the other two areas. In particular, administrative data systems covering the universe of recipients or workers record relatively few of these outcomes. We are thus left to use other data sources with much smaller samples and often no pre-CalWORKs data.

certificate data will be available. How useful it will be is less clear. Birth certificates contain names, but not SSNs. There is some prospect for probabilistically matching names to SSNs. It is not clear, however, where we would get a general file with names and SSNs on which to base the probabilistic match. Furthermore, even if we had such a link file; we remain concerned about the quality of the match. For now, we have decided not to pursue the third approach.

Table 4.4 summarizes the information on child and family well-being in our primary data sources. The large, statewide, administrative data systems have relatively little information once a family leaves aid. The CPS contains national data on a limited set of outcomes but does not identify county and has only small samples of current and recent recipients.

Table 4.4
Use of Various Data Sources--Child and Family Well-Being Outcomes

Specific Outcomes	CPS	6CHS	MEDS*	Q5*	6CWAD*	MEDS-* EDD
Household Resources and Poverty	X	X		X	X	
Marital Status	X	X		X	X	
Births, Their Marital Context, and Child Health	X	X				
Health Insurance	X	X	X	X	X	X
Foster Care, Child Abuse, and Child Living Arrangements	X	X				

Notes: X = The data contain this element (subject to quality assessment); * = Current recipients only.

To ameliorate this problem of lack of data, we are devoting approximately a third of project resources to the 6CHS, which will allow us to measure a range of otherwise unmeasurable outcomes. The 6CHS has (by design) information on essentially all the outcomes of interest. However, since there will be under 500 cases in each of the six focus counties (and nowhere else in the state or in other states) and both interviews will occur post-CalWORKs, these data will limit the possible analyses. In particular, our ability to evaluate the causal effect of CalWORKs on these outcomes using the 6CHS will be extremely limited. We discuss the 6CHS in more detail in Section 5.

Household Resources and Poverty

The previous section has considered earnings for recipients; however, earnings provide an incomplete depiction of household well-being. In particular, recipients may exit CalWORKs through marriage.

We will not know the identity of the new spouse, so we cannot find his earnings in the EDD data with which to compute total household resources. In addition, many current and recent CalWORKs households are eligible for large payments as part of the Earned Income Tax Credit (EITC).

Information on total household resources is available from several imperfect sources. CPS data have information for a small sample but do not identify county of residence, and the identification of recent welfare recipients is imperfect. The 6CHS will have information on total household resources for a sample in the 6 focus counties.

An alternative source is FTB information based on income tax returns. As such, this information includes nearly all sources of income, and it is possible to combine information for couples across two returns for married filing separately. We have begun exploring the availability of these data. Preliminary indications are promising. There is, however, a major concern. Tax filing is far from universal, especially in low-income populations; however, this situation should have improved considerably with the increased value of the EITC. We will use the 6CHS data to explore the completeness of tax filing and or ability to use EDD data to ameliorate the problems induced by non-filing.

Filing for the EITC is important in its own right. It is a major potential source of resources for current and recent CalWORKs households. Unusually high failure-to-file rates for households with earnings might suggest a lack of appropriate and effective guidance from county welfare department caseworkers to employed current and recent recipients. As of now, it seems that the FTB data will provide some descriptive post-reform information. It thus appears that, while some descriptive analysis of household poverty will be possible, a thorough analysis of the causal effects of CalWORKs on household income and poverty is unlikely.

Marital Status

The situation for marital status is similar to that for household resources and poverty. Information should be available in the CPS, the

6CHS, and the FTB files. Some descriptive analyses will be possible, but a thorough analysis of the causal effects of CalWORKs is unlikely due to the data limitations described above.

Births, Their Marital Context, and Child Health

Affecting births and nonmarital births was an explicitly stated goal of the PRWORA. For these outcomes, there is the potential for better data. The near-universal registration of births by birth certificate gives us excellent data on births and their marital context. Probabilistic matching by name with MEDS may allow us to separately analyze births to current recipients, recent recipients, and potential future recipients. Both descriptive and causal analyses (compared to AFDC/GAIN and across counties) would be possible.

In addition, the birth certificates also contain some information on health at birth. That information includes weight, congenital abnormalities, and prenatal care. In addition, information is available on whether Medi-Cal funded the birth. Again, both descriptive and causal analyses would be possible.

There is, however, a problem. It is not clear that the evaluation will have access to the confidential birth certificate data. Some earlier CDSS-REB studies have gained access to the data and the birth certificates were noted in the original CDSS RFP, but recently CDSS has not been able to obtain the data. If CDSS obtains the data, we will analyze it.

Health Insurance

Health insurance coverage is another measure of child and family well-being. We will have data on Medi-Cal coverage from the MEDS. Individuals receiving cash aid are presumptively eligible for Medi-Cal. Individuals leaving aid are eligible for cash aid for two years, and their children are eligible for longer if the household income is low enough. We will explore changes in Medi-Cal coverage using regression approaches.

Medi-Cal is not the only source of health insurance. As recipients move into the work place or marry, some of them will be covered by private health insurance. We will measure such coverage in the 6CHS and

in the CPS. These analyses will be primarily descriptive. They will help us to understand the Medi-Cal results.

Foster Care, Child Abuse, and Child Living Arrangements

There is considerable concern about the effect of PRWORA and CalWORKs on children. In particular, there is a fear that the work requirements will upset already fragile families, leading to child abuse and to the removal of children from their parents and to their placement in foster care. In addition, since time limits only apply to adults, there is a fear that children will be moved to the homes of other relatives to maintain the full benefit. This concern is real but perhaps less salient in California, where the child benefit continues.

Ideally, we would like to measure these outcomes and explore the effect of CalWORKs on them. Nominally, information on child neglect and foster care placements can be found in the Child Welfare Services/Case Management System (CWS/CMS). Those data appear to be less than ideal. First, the child welfare data system recently changed over from several county-specific data systems to a single CWS/CMS statewide. Second, the quality of the data collected during the transition is known to be poor, and the data will not be consistent over the transition. Together, these two data sources might allow a descriptive analysis. The Center for Social Services Research is currently processing these files for analysis under contract from CDSS. When they complete their processing of these files, we will confer with them about analysis strategies consistent with the structure and quality of the data.

FINANCIAL OUTCOMES

In addition to an analysis of outcomes--in the welfare system, in the labor market, and for children and families--the RFP requests a cost-benefit analysis. To perform such an analysis, we need to track costs. Much of the budget information comes from hard-copy sources. Here, we discuss briefly the principal data files describing expenditures that flow from CDSS to the counties.

Cash Aid Payments

The counties report detailed information on county aid payments by program category to CDSS on the CA237 and CA800 forms. The CA237 data are available in machine-readable form and we are already processing them for our caseload analyses. The CA800 form is submitted monthly. It does not include individual-level information. We discussed some of the issues in modeling the determinants of payments above. (See "Welfare System Outcomes".) As we noted there, county aid payments continue to be mandated under CalWORKs. Their level is not at county discretion. Beyond state legislation (e.g., regarding who is eligible, what is the benefit level, how much does the benefit decline with earnings and child support payments), the primary determinants of county aid payments are the level of the caseload, earnings, and sanction status. Each of these items will be explicitly considered in the analysis of welfare system outcomes--insofar as possible in all 58 counties, in the 19 Q5 counties, and in the 6 focus counties. Those analyses will serve as crucial input into the consideration of financial effects of CalWORKs.

County Administrative Expenditures

Cash aid payments are not the only county expenditures on CalWORKs recipients. One intended effect of PRWORA at the federal level and CalWORKs at the state level was to increase the resources available for services: case management and post-employment services; child care; alcohol, drug abuse, mental health treatment, and domestic violence; transitional health insurance; and (to a lesser degree than in the original GAIN program) education and training.

Counties report and claim reimbursement from CDSS for many of these expenditures through County Expense Reports. It appears that these reports summarize nearly all expenditures funded by CDSS and that some non-CDSS expenditures are also reimbursed through this mechanism. We have acquired and are processing the data for September 1992 to the present.

Other Data Sources for Financial Outcomes

The expenditures that flow through CDSS (and are recorded in these two data sources) are not the only financial outcomes (financial costs or financial benefits) relevant to CalWORKs. Other data sources will help us to gain a more complete picture of the financial effects of CalWORKs.

Among those data sources are the following. Total cash aid payments, by county, by month, by type of case (FG, UP, etc.) are reported on the CA237 forms discussed above. For those who file tax returns, information on tax payments and tax credits is available from the FTB data, also discussed above. Information on payroll taxes (Social Security and Medicare payroll taxes and, to a lesser extent, Unemployment Insurance contributions) paid can be inferred from the FTB data. Federal flows to California are available in federal documents.

Other federal, state, county, and local governments and agencies also have financial costs and benefits from CalWORKs. We are still exploring other sources of information on those financial effects. It appears that they will primarily be available as paper records. As such, they are likely to provide less detail and, therefore, are likely to support only less detailed analyses. In as much as detailed information (usually in computerized form) is available, we will need to consider whether the additional information that might be gained is worth the likely large fixed costs of processing and understanding the data.

Currently, it is our sense that the primary cost impacts are on CDSS and the county welfare departments. In particular, much of the financial impacts appear to involve federal and state maintenance of effort requirements and the details of provisions for carry-forward of unexpended funds. These issues can be explored using the two primary CDSS data systems.

In contrast, non-CDSS financial effects are harder to measure and appear to be less salient. Therefore, we have given less attention and fewer resources to the acquisition, processing, and understanding of financial information from other agencies. Our analyses of the financial effects are ongoing. As our understanding of these financial

effects improves, and, as appropriate, we will revisit these resource-allocation decisions.

5. THE SIX COUNTY HOUSEHOLD SURVEY

As we noted in the previous section, while the administrative data provide good to outstanding information on many welfare systems and labor market outcomes, the data are considerably weaker in providing information on child and family well-being. To describe outcomes not recorded in the administrative records, the Statewide CalWORKs Evaluation is devoting approximately a third of its resources to the 6CHS, a new household survey effort.

This section provides a more detailed description of that effort. We begin with a broad overview of the design and a detailed discussion of the sampling plan. We then discuss the strengths and weaknesses of the design. Finally, we review the content of the survey.

OVERVIEW OF DESIGN

Table 5.1 summarizes the key features of the 6CHS. The survey is being fielded in the six focus counties specified by CDSS in its RFP: Alameda, Butte, Fresno, Los Angeles, Sacramento, and San Diego. In each of the six focus counties, the survey will interview approximately 475 current and recent welfare recipients, under current assumptions about response rates.

Due to cost considerations, interviews will occur only in English and Spanish.²² Cases recorded as speaking any other language in the county files will be excluded from the sampling frame. Cases encountered in the field where no adult speaks either language will be excluded at that time.

The sample will be drawn based on the most recent MEDS file available when the sample needs to be drawn, approximately three months before interviewing begins. At that time, the MEDS data will be approximately one month old. Thus, for example, for interviews to be

²² While non-trivial shares of the caseload speak only some other language, no other single language is spoken by a large enough share of the caseload. Furthermore, the fixed cost of formulating the survey instrument into another language and then hiring interviewers and supervisors for another language is quite high.

conducted in January, the sample will need to be drawn late in October, based on the September MEDS file. As noted earlier, in addition, it appears that new cases are sometimes reported with a lag.

Table 5.1
Key Design Features of the 6CHS

Panel study fielded in the six focus counties
One hour in-person interviews in Winter/Spring 2000
Telephone follow-up approximately one year later
Current and former recipients of FG and UP and Child-Only cash aid grants, with an oversample of UP (two-parent) cases
Approximate sample size of 475 in each county (under current assumptions about response rates)
Interviews conducted in English and Spanish
Interviewing one adult (over age 18) per sampled assistance unit
Some geographic clustering within the counties to reduce survey costs
Limitations:
- Sample size
- Representativeness of the follow-up sample
- Statewide generalizability
- Only current and recent recipients
- Only post CalWORKs
- Only in California

The sample will have two strata in each county. We anticipate interviewing approximately 325 one-parent households and 150 two-parent households. Since statewide, two-parent households represent approximately 17 percent of the cases, the implied sampling rates are twice as high for two-parent households as for one-parent households. This oversampling is included for two reasons. First, because the two-parent sample is smaller, larger samples are required to make statements about this program. Second, the recent decision of the State of California to establish a Separate State Program (SSP) for two-parent cases (see ACL-54, August 12, 1999) has focused additional substantive attention on this subgroup. In practice, we will select the oversample based on the FG/UP distinction on the administrative data files. The two concepts (one-parent/two-parent versus FG/UP) are similar but not

identical. Our sampling approach is determined by the available data elements as of the time we need to select our sample.

Within the strata, within each of the six focus counties, we will select a clustered random sample. Anyone who received cash aid in the 12 months immediately preceding the selection of the sample will have an equal probability of selection.²³ By design, this sample will include those continuously on aid through this period, those who entered aid during this period, those who left aid during this period, and those who have exited and reentered aid over this period. We note that studies of recent leavers have received considerable attention recently (Loprest, 1999) and are likely to be of considerable interest in California as well.

To lower field costs, the sample will be geographically clustered. We are currently finalizing the geographical clustering scheme. Our basic approach is based on zip codes in the MEDS file which we will use to select the sample. We will create geographical cluster based on most recent recorded zip code of residence. We will then select zip codes with probability proportional to size and select samples of fixed size within each zip code. Given this sampling scheme, we will create appropriate weights for the sampling scheme.

This is our general sampling plan. Ideally, each cluster should have enough cases to require about a third of an interviewer's time. We project nine to ten interviewers per county, so we need to select approximately 30 clusters. However, this approach is not feasible without some adjustment. Some zip codes have too few welfare cases to support a third of an interviewer, while some zip codes will have so many cases that they could support a third of an interviewer several times over. Thus, we might want to at least allow for the possibility that we would want to assign multiple interviewers to a zip code. We are currently developing methods to handle (i.e., group or split) zip codes in these extreme cases.

²³ We note, but will ignore, the possibility that an individual could be selected for the sample more than once. This would happen if within the twelve-month window the individual had received cash aid in more than one focus county.

Table 5.2
Tabulations for Sample Characteristics

	Alameda	Butte	Fresno	Los Angeles	Sacramento	San Diego
On aid in the 1st month of the 12-month sample window	81.6	76.3	77.4	81.5	80.0	81.0
Continuously on aid during the sample window	46.2	35.7	40.5	52.0	51.1	42.7
Intermittently on aid in the sample window	16.2	25.7	16.7	10.4	10.5	10.6
On aid at the baseline interview	65.5	61.5	60.8	67.1	67.7	57.1
Continuously on aid between the surveys	32.9	25.1	31.5	43.5	43.0	30.0
Intermittently on aid between the surveys	13.9	15.6	10.4	4.6	5.9	5.6
On aid at the follow-up interview	45.9	38.9	42.0	49.7	49.4	36.6
UP caseload	14.1	23.8	28.6	18.2	24.4	19.5
Race						
White	20.0	78.3	21.6	18.1	45.2	34.4
Hispanic	11.8	7.5	50.8	45.2	13.3	34.8
Black	52.1	3.6	12.3	29.3	25.0	20.5
Other	16.1	10.6	15.3	7.4	16.5	10.3
Language						
English	83.3	92.2	83.5	72.6	80.2	76.9
Spanish	3.0	2.8	11.1	18.1	2.5	15.2
Other	13.7	5.0	5.4	9.3	17.3	7.9

Notes: These simulations were generated using the MEDS data, defining the sample window as March 1997 to February 1998, May 1998 as the date of the baseline interview, and May 1999 as the date of the follow-up interview.

Table 5.2 gives some summary statistics from some projections of the household survey sample characteristics based on recent (as of the writing of this document) experience. Like the final sample, the estimates are constructed from the MEDS file. We applied the final sampling rules to the most recent file (August 1999). The final sample

will be built in the same way, but from a later file and including geographical clustering.

Selected individuals--both current and recent welfare recipients--will be interviewed twice: first in the January 2000-May 2000 timeframe, and then again in the January 2000-April 2001 timeframe. The first line of the table indicates that approximately 20 percent of the sample will have come on aid since the beginning of the sample window. About half the sample will have been on aid continuously in the sample window. Roughly 12 percent will have had multiple spells of aid receipt; the questionnaire is designed to ask about the most recent spell of aid. In the baseline survey, we expect about 35 percent of the sample to be welfare "leavers," not on aid at the time of the interview. By the time of the follow-up survey, we expect 53 percent of the sample to be leavers. Note that there will be no "refresh sample;" thus, we will have no observations on those who enter aid for the first time after the initial sample is drawn in late 1999/early 2000.

LIMITATIONS OF THE DESIGN

This sampling plan will allow us to describe outcomes for child well-being that are otherwise poorly measured. In particular, the design is specified to provide information on child and family well-being, for those still on aid and for those who have left. These outcomes figured prominently in the debate leading up to TANF and CalWORKs and are likely to receive considerable weight in judgements about the success or failure of the reforms. We believe that the 6CHS will allow us to describe these outcomes under CalWORKs and those descriptions are a key part of our evaluation.

The design, however, has important limitations that should be noted. First, both of the interviews will take place after the passage and early implementation of CalWORKs. Thus, the 6CHS will be of only limited usefulness in comparing CalWORKs outcomes to outcomes prior to CalWORKs or to what outcomes would have been if the old AFDC program had been left in place.

Second, the 6CHS will be fielded only in the six focus counties. This design is consistent with the spirit of CDSS-REB's RFP, but it has

limitations. Since there are no observations outside of California, the 6CHS's utility for interstate comparisons (e.g., what California's outcomes would have been if it had adopted a TANF program more closely resembling that of other states) will be limited. Probably more salient, however, is a related concern. As implied by its name, the 6CHS is to be fielded only in six of California's counties. Together, these counties contain about half the state's population and about half of the state's welfare caseload. Nevertheless, these counties are not representative of the state as a whole. The counties were purposively chosen by CDSS for the evaluation. They are skewed toward the larger, more urban counties, and the state's smallest counties are not represented at all.

Analyses will sometimes explore results for each county separately. However, the sample sizes will not be large for many such analyses, so cross-county comparisons will often not be able to detect statistically significant differences. Instead, analyses of these data will often pool the results across the six counties and report results for the CalWORKs program as a whole. This is our analysis plan. Implicit in reporting results, this approach either assumes that outcomes are common across the state or gives equal weight to the six focus counties and no weight to the other 52 counties. Clearly, neither of these alternatives is correct. However, any other alternative (e.g., adding more counties, perhaps randomly chosen, or adding a sample of individuals randomly allocated throughout the remaining 52 counties) would either be inconsistent with the spirit of the RFP and its specification of focus counties or would be prohibitively expensive (or both).

SURVEY CONTENT

The content of the survey is focused to complement the existing administrative data and case files. Table 5.3 gives the titles of the main sections of the survey. Copies of the full instrument will be available on the project web site, listed in the Preface.

Table 5.3
Sections of the Survey

Section	Title
	Introduction
A	Welfare
B	Educational History
C	Household Roster, Demographics, and Household Composition
D	Employment
E	Transportation
F	Spouse/Partner Proxy Employment and Health Insurance
G	Household Income
H	Other Assistance, Financial Hardship, and Food Security
I	Housing Information
J	Child Care and Child Education
K	Child Support and Contact with Absent Parents
L	Health and Behavioral Health
M	Closing Section

The survey begins with an informed consent statement. That statement informs recipients that the survey is being sponsored by CDSS, that CDSS will receive the raw data, including identifiers, but that CDSS promises to use the data only for research purposes. It also informs them that their responses will be linked to their administrative data, that RAND will resurvey them once, and that CDSS may recontact them later. Finally, it notes that participation is voluntary. It is not required by law. Failure to respond will not affect welfare payment or any such benefits. It then asks if they are willing to be interviewed. A standard screener follows.

The first substantive section of the survey concerns the welfare system. We will know the history of aid receipt from the administrative records, so there are few objective questions, generally those that appear to be recorded poorly in the administrative data: current program requirements and compliance status, participation in Job Club, experience with sanctions, receipt of support services. There are also several subjective questions: reasons for entering aid, knowledge of program rules (including time limits, work requirements, family cap, transitional MediCal), and attitudes toward activities and caseworkers.

Several sections then collect background information. Section B collects information on schooling. Section C collects information on

basic demographics and household composition, including the location of all own children (even outside the household) and location of all parents of children in the household.

Section D considers employment. We have detailed longitudinal histories of employment and earnings from the MEDS-EDD match. Here, we collect some objective information not available in the MEDS-EDD data (hours, hourly wage, schedule, industry, occupation, type of employer, on-the-job training, and health insurance from jobs) and some subjective information (why not working, reason for quitting a job).

Section E collects information on transportation: usual method of travelling to activities, time required, and receipt of subsidy.

Section F asks about employment and health insurance of a partner (whether or not there is a formal marriage). Section G attempts to identify other sources of household income for the past month beyond welfare and own and spouse labor earnings. It also includes a brief battery of questions on assets, including automobile ownership.

The survey then turns to the child and family well-being issues that are not well recorded in the administrative data. Section H includes some questions on social support, financial hardship, and part of the now standard battery on "food security" drawn from Work Pays (which will allow some pre-/post-CalWORKs comparisons). Section I collects information on housing, including some questions on "housing security" (also from Work Pays). Section J collects some information on child care and child education. Section K collects information on child support arrangements, receipt, and contact with the absent parent. Section L collects some information on own and child health and behavioral health, including limited batteries on mental health status and alcohol use. The survey concludes by collecting contact information to help in locating respondents at the second wave of the survey.

6. PROJECT FOCUS AND STATUS

This report has described RAND's plans for the impact analysis component of the CDSS-funded Statewide CalWORKs Evaluation. In particular, it has explored what we want to know, what methods we will use, and what data are available. In this section, we discuss the focus of analyses and the status of these plans.

FOCUS OF ANALYSES

Considerable data resources exist for the evaluation. For some outcomes, they provide large samples (even the universe), for long periods, with detailed background and outcome information, in easily accessible and analyzable form. For other outcomes, the available data are more limited. Clearly, what outcomes can be described and what net causal effects can be estimated will be limited by the available data.

The State of California has set aside considerable funds for the Statewide CalWORKs Evaluation. However, the scope of the changes required by the CalWORKs legislation and thus the possible set of outcomes and impacts to study is also large. While we propose some primary data collection--the 6CHS--that approach is expensive per case. In addition, processing each new secondary data source requires large fixed costs, which include negotiating for access to data, processing the data into an easily analyzable file format, and working with those who produced the data to understand what is reliable and what the responses mean. Despite the generous funding, choices will need to be made and priorities will need to be set.

The previous section reviewed the substantive outcome domains, the available data, and the possible analyses. Here, we list what we will do.

1. Caseload, Aid Payments, and Costs

Caseloads and aid payments are the immediate outcomes of welfare programs. The CA237 files provide a long time series of aggregate data on both caseloads and aid payments. For caseloads, the MEDS data provide disaggregated data allowing analyses by demographic

subpopulations, program type (AFDC versus UP versus child-only versus Medi-Cal only) and dynamic characteristics (e.g., spell length, total accumulated months on aid, age at first receipt). Therefore, for aggregate aid payments and disaggregated caseloads, we expect to estimate the causal effect of CalWORKs programs for every county in the state. For caseloads, we will also perform dynamic analyses. We will use both regression approaches and matching approaches. We will compare caseload and aid payments under CalWORKs to what these outcomes would have been under each of the three baselines--if AFDC had continued, across counties, and (for aggregate data) across states. For caseloads, we will also do dynamic analyses. In addition, that 6CHS will give us some subjective information on why recipients enter, leave, and return to aid.

Corresponding to these effects of CalWORKs on the caseload are its effects on costs. To move almost all the caseload to work and self-sufficiency within time limits, CalWORKs appears to have been envisioned as a program providing more intensive services per case than GAIN. The County Expense Reports and other sources of data on cash aid payments will provide detailed information on those expenditures. Understanding them is one of our highest priorities.

The combination of caseload declines, per-case cash-aid declines, and maintenance-of-effort requirements appear to mean that the crucial issues are not the total levels of expenditures. Instead, two other issues appear to be important. The first issue concerns the allocation of expenditures. State maintenance-of-effort requirements imply that counties must spend some funds. Counties have some discretion on how to spend these funds. We want to understand those choices, why they were made, and their implications.

The second issue concerns carry-forwards. The individual state-funding streams have varying requirements about when and how the monies must be spent. In particular, some of the funds can be carried over from one year to the next. Preliminary indications are that considerable funds are indeed being carried over. Better understanding the source, magnitude, and motivation for these carry-overs is a major goal of our evaluation.

2. Program Activities

PRWORA and CalWORKs specify a sequence of activities and procedures for dealing with noncompliance. Understanding how individual participants flow through the system is crucial to characterizing county CalWORKs programs and to understanding their effects on recipients. As noted in Section 4, we are currently exploring the extent to which such information on program activities can be extracted from the administrative data systems for the six focus counties. Preliminary investigations suggest that processing these data is likely to require close collaboration with the county coordinators and considerable RAND staff resources, leading in the end to valuable but incomplete data that will vary in content and quality across the six counties. Despite these cost and content concerns, we have allocated a large fraction of project resources to this task. We expect this effort to yield disaggregated, cross-sectional, and dynamic descriptions of program activities. The results will be crucial to the process analysis, the impact analysis, and the cost-benefit analysis. In addition, the 6CHS collects some information on program activities and also information on subjective experiences with program activities and contact with caseworkers.

We understand that there is considerable interest in estimating the causal effect of individual program components and legal requirements on program outcomes. We will devote some resources to these questions and expect to produce some results. Considerable caution, however, is indicated about the scope of the results and their value. This is one of the focuses of the ongoing methodological work.

3. Employment and Earnings

The CalWORKs model involves moving people off of aid through employment. Employment is likely to be a key component of the federal and state participation requirements, and state incentive payments to the counties are based on exits resulting from employment. Finally, employment and earnings are key determinants of household income, poverty, and child and family well-being. For current and former welfare recipients, the MEDS-EDD match provides excellent disaggregated data. Therefore, for disaggregated employment and earnings, we expect

to do statewide estimation of the causal effect of CalWORKs programs. We will consider both cross-sectional measures (e.g., status at a point in time) and dynamic measures (e.g., whether people are finding work faster after first receipt of aid, whether earnings are increasing with job tenure). We will use both regression approaches and matching approaches. We will compare employment and earnings under CalWORKs to what these outcomes would have been under two baselines--if AFDC had continued, and across counties. In particular, our analysis will track progress toward quarterly earnings consistent with self-sufficiency, defined both relative to the poverty line when combined with other transfer programs and relative to becoming ineligible for cash aid.

Hours worked is a key component in federal and state participation requirements. However, the EDD data do not report hours. As we discussed in Section 4, limited data (in terms of period, counties, quality, and sample size) are/will be available. While we will do some analyses of hours of work using the 6CHS, the data are not sufficient to allow a detailed analysis.

In addition to moving current recipients to self-sufficiency, PRWORA and CalWORKs are intended to change the life courses of potential future recipients. While such analyses are potentially important, the data issues appear daunting. Therefore, at least temporarily, we are deferring such analyses.

4. Child and Family Well-Being

Ultimately, the success of CalWORKs will be judged by balancing changes in caseloads and program cost against its effects on children and families. As we discussed in Section 4, however, data on child and family well-being are weaker than data on welfare system and self-sufficiency outcomes. With several noted exceptions, ideal data--consistent across the pre- and post-CalWORKs periods, covering the entire state (or even a large number of counties), with large samples, and consistent data definitions--are not available. Thus, despite the importance of the outcomes, full analyses of the effects of CalWORKs will not be possible. Furthermore, as noted, processing and analyzing

each additional data set have high fixed costs. Thus, given all this, our current plans are as follows.

Birth certificates provide high-quality data on several crucial outcomes, including out-of-wedlock births, health care received during pregnancy, and health of newborns at birth. If we can get access to these data, we will analyze them. As of now, there is considerable doubt about whether access will be granted.

FTB data provide information on household income and marital status for those who file taxes. These data provide an important potential complement to the MEDS-EDD match. The information, however, is only available for those filing tax returns. With the new higher EITC, tax-filing rates appear to have increased. We will carefully explore the selectivity of tax filing and therefore the utility of FTB data.

Information on other outcomes is much more limited. To address these weaknesses, we will analyze national CPS data and devote nearly a third of the evaluation resources to the fielding of a major new survey effort, the 6CHS. We will provide descriptive tabulations of poverty, family structure, and health insurance coverage from the CPS and the 6CHS.

Finally, we will collect aggregate data from other sources (e.g., foster care, child abuse, crime, and contact with the criminal justice system). Considerations about data availability, data comparability across time, and the fixed costs of processing more data currently lead us now to lean against processing and analyzing the disaggregated data.

STATUS

This report has described the current status of RAND's planning for the impact analysis component of the CDSS-funded Statewide CalWORKs Evaluation. RAND's response to CDSS's RFP sketched an analysis plan. Limited time, page limits, and uncertainty about the data systems required that this earlier discussion be incomplete.

RAND was awarded the evaluation contract in October 1998. Since then, RAND staff has been working to specify more completely a plan for the impact analysis. We have begun the process of acquiring and processing data systems. In some cases, we have begun preliminary

analysis. Our methodological work on matching approaches to estimating causal effects is proceeding.

By contract, the first impact analysis report is to be released in October 2000. Allowing sufficient time for CDSS review, RAND revisions, and printing, a draft will need to be forwarded to CDSS, the counties and other stakeholders in early August. Given our current schedule, that report is likely to contain:

- Descriptive analyses of program activities (from GAIN25, WTW25, and 6CWAD);
- Descriptive and causal analyses of caseloads (from MEDS), aid payments (statewide from CA 237 and in the six Focus Counties from 6CWAD);
- Descriptive and causal analyses of earnings and employment for current and recent recipients (from the matched MEDS-EDD file);
- Budget analyses (from CA 237 and the County Expense Form);
- Validation of our methods, using data from the GAIN experiments of the early-1990s.

In addition to updating each of these results, the second and final impact analysis report due in October 2001 will include:

- Descriptive analyses of child and family well-being outcomes (from both waves of the 6CHS and the CPS);
- Descriptive analyses of fertility and health of new-borns (from California birth certificates, if we can get access to the data);
- Descriptive analyses of child endangerment, foster care and child living arrangement (from CWS/CMS--if we can get access to cleaned data--and from the CPS);
- Descriptive analyses of family income and EITC filing (from FTB tabulations).

This is our current plan. Our data acquisition, processing, and analysis work are at varying stages for varying data sets. While some of the data uncertainty has been resolved, much remains.

Our analysis plan will evolve as we learn more about the data and as preliminary results emerge. We expect our plans for the impact analysis to continue to evolve over the remaining two years of the evaluation. Future quarterly progress reports, meetings of the Advisory Committee, draft documents, and presentations of plans and results before academic and policy audiences will provide opportunities for RAND to share these evolving plans with CDSS and the broader research community. Feedback from future written and oral presentations will also help RAND improve the technical quality of its analyses and the allocation of available resources to the tasks of greatest interest to CDSS.

REFERENCES

- Blank, R. (1997). "What Causes Public Assistance Caseloads to Grow?" National Bureau of Economic Research Working Paper No. 6343.
- California Department of Social Services (CDSS) (1998). Request for Proposal (RFP) 98-02: Statewide Evaluation of the CalWORKs Program, Sacramento, CA.
- Council of Economic Advisers (1997). "Technical Report: Explaining the Decline in Welfare Receipt, 1993-96." A Report by the Council of Economic Advisers.
- Council of Economic Advisers (1999). "Economic Expansion, Welfare Reform, and the Decline in Welfare Caseloads: An Update." A Technical Report by the Council of Economic Advisers.
- Dehejia, R. and S. Wahba (1999). "Causal Effects in Non-experimental Studies: Re-evaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94(448):1053-1062.
- Ebener, P. and J. A. Klerman (1999). *Welfare Reform in California: Results of the 1998 All-County Implementation Survey*, MR-1052-CDSS, Santa Monica, CA: RAND.
- Ebener, P., E. Roth, and J. A. Klerman (1999b). *Welfare Reform in California: Results of the 1998 All-County Implementation Survey, Appendix*, MR-1052/1-CDSS, Santa Monica, CA: RAND.
- Figlio, D. and J. Ziliak (1999). "Welfare Reform, the Business Cycle, and the Decline in AFDC Caseloads." March manuscript.
- Friedlander, D., D. H. Greenberg, and P. K. Robins (1997). "Evaluating Government Training Programs for the Economically Disadvantaged." *Journal of Economic Literature*, 35(4):1809-1855.
- Haider, S., J. A. Klerman, J. M. Hanley, L. McDonald, E. Roth, L. Hiatt, and M. Suttorp (1999). *Welfare Reform in California: Design of the Impact Analysis: Preliminary Investigations of Caseload Data*, PM-940-CDSS, Santa Monica, CA: RAND.
- Heckman, J. J. and V. J. Hotz (1989). "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training (in Applications and Case Studies)," *Journal of the American Statistical Association*, 84(408):862-874; 875-880.

- Hill, C. J., V. J. Hotz, C. Mullin, and J. K. Scholz (1997). "EITC Eligibility, Participation and Compliance Rates for AFDC Households: Evidence from the California Caseload."
- Hotz, V. J., G. Imbens, and J. Mortimer (1999). "Predicting the Efficacy of Future Training Programs Using Past Experiences," Technical Working Paper 238, Cambridge, MA: National Bureau of Economic Research, <http://www.nber.org/papers/T0238>.
- Hotz, V. J. (1992). "Designing an Evaluation of the Job Training Partnership Act," in Charles F. Manski and Irwin Garfinkel (eds.), *Evaluating Welfare and Training Programs*, Cambridge, MA and London: Harvard University Press, pp. 76-114.
- Hoynes, H. (1996). "Local Labor Markets and Welfare Spells: Do Demand Conditions Matter?" NBER Working Paper 5643.
- Klerman, J. A., E. Reardon, and P. Steinberg (1998). *RAND Statewide CalWORKs Evaluation: An Overview*, DB-252-CDSS, Santa Monica, CA: RAND.
- Klerman, J. A., E. Reardon, P. A. Ebener, and J. Hawes-Dawson (1999). *Welfare Reform in California: Design of the Impact Analysis: Six-County Household Survey*, PM-941-CDSS, Santa Monica, CA: RAND.
- Lalonde, R. J. (1986). "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *The American Economic Review*, 76(4):604-620.
- Loprest, P. (1999). *How Families That Left Welfare Are Doing: A National Picture*, New Federalism Issues and Options for the States, Paper No. B-1, Washington, D.C.: The Urban Institute.
- MacCurdy, T. and M. O'Brien-Strain (1997). *Who Will Be Affected by Welfare Reform in California?* San Francisco, CA: Public Policy Institute of California.
- Martini, A. and M. Wiseman (1997). "Explaining the Recent Decline in Welfare Caseloads: Is the Council of Economic Advisers Right?" Mimeograph, July, Washington, D.C.: The Urban Institute.
- Meyer, B. (1995). "Natural and Quasi-Experiments in Economics," *Journal of Business and Economic Statistics*, 12(2):151-161.
- Moffitt, R. (1999). "The Effect of Pre-PRWORA Waivers on AFDC Caseloads and Female Earnings, Income, and Labor Force Behavior." May manuscript.
- Rosenbaum, P. and D. Rubin (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70(1):41-55.

Wallace and Blank (1999). "What Goes Up Must Come Down? Explaining Recent Changes in Public Assistance Caseloads," February manuscript.

Zellman, G., J. Klerman, E. Reardon, and P. Steinberg (1999b). *Welfare Reform in California: State and County Implementation of CalWORKs in the First Year, Executive Summary*, MR-1051/1-CDSS, Santa Monica, CA: RAND.

Zellman, G., J. Klerman, E. Reardon, D. Farley, N. Humphrey, T. Chun, and P. Steinberg (1999a). *Welfare Reform in California: State and County Implementation in the First Year*, MR-1051-CDSS, Santa Monica, CA: RAND.

Ziliak, J., D. Figlio, E. Davis, and L. Connolly (1997). "Accounting for the Decline in AFDC Caseloads: Welfare Reform or Economic Growth?" Institute for Research on Poverty Discussion Paper No. 1151-97.